(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016506 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/27143

(22) International Filing Date: 15 August 2002 (15.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/313,249	17 August 2001 (17.08.2001)	US
60/314,752	24 August 2001 (24.08.2001)	US
60/317,818	7 September 2001 (07.09.2001)	US
60/317,824	7 September 2001 (07.09.2001)	US
60/324,040	21 September 2001 (21.09.2001)	US
60/324,586	24 September 2001 (24.09.2001)	US
60/343,980	2 November 2001 (02.11.2001)	US
60/334,229	28 November 2001 (28.11.2001)	US
60/357,002	13 February 2002 (13.02.2002)	US
60/362,439	6 March 2002 (06.03.2002)	US
60/366,041	19 March 2002 (19.03.2002)	US
60/376,988	30 April 2002 (30.04.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WARREN, Bridget, A.** [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **HONCHELL, Cynthia, D.** [US/US]; 158 Laurel Street, San Carlos, CA 94070 (US). **LEHR-MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **YUE, Huibin** [US/US]; 1170 South Stelling Road, Cupertino, CA 95014 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **SANJANWALA, Bharati** [IN/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **NGUYEN, Danniel, B.** [US/US]; 4922 Arnica Court, San Jose, CA 95111 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **ASTROMOFF, Anna** [US/US]; 132 Winding Way, San Carlos, CA 94070 (US). **DING, Li** [CN/US]; 11036 Ridgecorde Lane, Creve Coeur, MO 63141 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **JIN, Pei** [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). **FU, Glenn, K.** [US/US]; 4215 Roscommon Way, Dublin, CA 94568 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: SECRETED PROTEINS

(57) Abstract: Various embodiments of the invention provide human secreted proteins(SECP) and polynucleotides which identify and encode SECP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of SECP.

WO 03/016506 A2

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SECRETED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, secreted proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and secreted proteins.

BACKGROUND OF THE INVENTION

Protein transport and secretion are essential for cellular function. Protein transport is mediated by a signal peptide located at the amino terminus of the protein to be transported or secreted. The signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent protein from the ribosome to a particular membrane bound compartment such as the endoplasmic reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that transit through the secretory pathway are either secreted into the extracellular space or retained in the plasma membrane. Proteins that are retained in the plasma membrane contain one or more transmembrane domains, each comprised of about 20 hydrophobic amino acid residues. Secreted proteins are generally synthesized as inactive precursors that are activated by post-translational processing events during transit through the secretory pathway. Such events include glycosylation, proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur during protein transport include chaperone-dependent unfolding and folding of the nascent protein and interaction of the protein with a receptor or pore complex. Examples of secreted proteins with amino terminal signal peptides are discussed below and include proteins with important roles in cell-to-cell signaling. Such proteins include transmembrane receptors and cell surface markers, extracellular matrix molecules, cytokines, hormones, growth and differentiation factors, enzymes, neuropeptides, vasomediators, cell surface markers, and antigen recognition molecules. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 557-560, 582-592.)

Cell surface markers include cell surface antigens identified on leukocytic cells of the immune system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of differentiation" based on common immunocytochemical localization patterns in various

differentiated and undifferentiated leukocytic cell types. Antigens in a given cluster are presumed to identify a single cell surface protein and are assigned a "cluster of differentiation" or "CD" designation. Some of the genes encoding proteins identified by CD antigens have been cloned and verified by standard molecular biology techniques. CD antigens have been characterized as both transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A.N. et al. (1995) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 17-20.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, MPs may be heavily glycosylated and may contain an Arginine-Glycine-Aspartate (RGD) tripeptide motif which may play a role in adhesive interactions. MPs include extracellular proteins such as fibronectin, collagen, galectin, vitronectin and its proteolytic derivative somatomedin B; and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) Kidney Int. 51:1413-1417; Sjaastad, M.D. and W.J. Nelson (1997) BioEssays 19:47-55.)

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W. et al. (1997) J. Biol. Chem. 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W. et al. (1993) J. Biol. Chem. 268:5879-5885). Hemomucin is a novel *Drosophila* surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U. et al. (1996) J. Biol. Chem. 217:12708-12715).

Tuftelins are one of four different enamel matrix proteins that have been identified so far. The other three known enamel matrix proteins are the amelogenins, enamelin and ameloblastin. Assembly of the enamel extracellular matrix from these component proteins is believed to be critical in producing a matrix competent to undergo mineral replacement (Paine, C.T. et al. (1998) Connect Tissue Res. 38:257-267). Tuftelin mRNA has been found to be expressed in human ameloblastoma

tumor, a non-mineralized odontogenic tumor (Deutsch, D. et al. (1998) Connect. Tissue Res. 39:177-184).

Olfactomedin-related proteins are extracellular matrix, secreted glycoproteins with conserved C-terminal motifs. They are expressed in a wide variety of tissues and in a broad range of species, from *Caenorhabditis elegans* to *Homo sapiens*. Olfactomedin-related proteins comprise a gene family with at least 5 family members in humans. One of the five, TIGR/myocilin protein, is expressed in the eye and is associated with the pathogenesis of glaucoma (Kulkarni, N.H. et al. (2000) Genet. Res. 76:41-50). Research by Yokoyama, M. et al. (1996; DNA Res. 3:311-320) found a 135-amino acid protein, termed AMY, having 96% sequence identity with rat neuronal olfactomedin-releated ER localized protein in a neuroblastoma cell line cDNA library, suggesting an essential role for AMY in nerve tissue. Neuron-specific olfactomedin-related glycoproteins isolated from rat brain cDNA libraries show strong sequence similarity with olfactomedin. This similarity is suggestive of a matrix-related function of these glycoproteins in neurons and neurosecretory cells (Danielson, P.E. et al. (1994) J. Neurosci. Res. 38:468-478).

Mac-2 binding protein is a 90-kD serum protein (90K), a secreted glycoprotein isolated from both the human breast carcinoma cell line SK-BR-3, and human breast milk. It specifically binds to a human macrophage-associated lectin, Mac-2. Structurally, the mature protein is 567 amino acids in length and is proceeded by an 18-amino acid leader. There are 16 cysteines and seven potential N-linked glycosylation sites. The first 106 amino acids represent a domain very similar to an ancient protein superfamily defined by a macrophage scavenger receptor cysteine-rich domain (Koths, K. et al. (1993) J. Biol. Chem. 268:14245-14249). 90K is elevated in the serum of subpopulations of AIDS patients and is expressed at varying levels in primary tumor samples and tumor cell lines. Ullrich, A. et al. (1994; J. Biol. Chem. 269:18401-18407) have demonstrated that 90K stimulates host defense systems and can induce interleukin-2 secretion. This immune stimulation is proposed to be a result of oncogenic transformation, viral infection or pathogenic invasion (Ullrich et al., *supra*).

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94). Plexins are neuronal cell surface molecules that mediate cell adhesion via a homophilic binding mechanism in the presence of calcium ions. Plexins have been shown to be expressed in the receptors and neurons of

particular sensory systems (Ohta, K. et al. (1995) Cell 14:1189-1199). There is evidence that suggests that some plexins function to control motor and CNS axon guidance in the developing nervous system. Plexins, which themselves contain complete semaphorin domains, may be both the ancestors of classical semaphorins and binding partners for semaphorins (Winberg, M.L. et al (1998) Cell 95:903-916).

Human pregnancy-specific beta 1-glycoprotein (PSG) is a family of closely related glycoproteins of molecular weights of 72 KDa, 64KDa, 62KDa, and 54KDa. Together with the carcinoembryonic antigen, they comprise a subfamily within the immunoglobulin superfamily (Plouzek, C.A. and J.Y. Chou, (1991) Endocrinology 129:950-958) Different subpopulations of PSG have been found to be produced by the trophoblasts of the human placenta, and the amnionic and chorionic membranes (Plouzek, C.A. et al. (1993) Placenta 14:277-285).

Torsion dystonia is an autosomal dominant movement disorder consisting of involuntary muscular contractions. The disorder has been linked to a 3-base pair mutation in the DYT-1 gene, which encodes torsin A (Ozelius, L.J. et al. (1997) Nat. Genet. 17:40-48). Torsin A bears significant homology to the Hsp100/Clp family of ATPase chaperones, which are conserved in humans, rats, mice, and *C. elegans*. Strong expression of DYT-1 in neuronal processes indicates a potential role for torsins in synaptic communication (Kustedjo, K. et al. (2000) J. Biol. Chem. 275:27933-27939 and Konakova M. et al. (2001) Arch. Neurol. 58:921-927).

Autocrine motility factor (AMF) is one of the motility cytokines regulating tumor cell migration; therefore identification of the signaling pathway coupled with it has critical importance. Autocrine motility factor receptor (AMFR) expression has been found to be associated with tumor progression in thymoma (Ohta Y. et al. (2000) Int. J. Oncol. 17:259-264). AMFR is a cell surface glycoprotein of molecular weight 78KDa.

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category includes small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category includes hydrophilic hormones that function by binding to cell surface receptors that transduce signals across the plasma membrane. Examples of such hormones include amino acid derivatives such as catecholamines (epinephrine, norepinephrine) and histamine, and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al.

(1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Pro-opiomelanocortin (POMC) is the precursor polypeptide of corticotropin (ACTH), a hormone synthesized by the anterior pituitary gland, which functions in the stimulation of the adrenal cortex. POMC is also the precursor polypeptide of the hormone beta-lipotropin (beta-LPH). Each hormone includes smaller peptides with distinct biological activities: alpha-melanotropin (alpha-MSH) and corticotropin-like intermediate lobe peptide (CLIP) are formed from ACTH; gamma-lipotropin (gamma-LPH) and beta-endorphin are peptide components of beta-LPH; while beta-MSH is contained within gamma-LPH. Adrenal insufficiency due to ACTH deficiency, resulting from a genetic mutation in exons 2 and 3 of POMC results in an endocrine disorder characterized by early-onset obesity, adrenal insufficiency, and red hair pigmentation (Chretien, M. et al. (1979) Can. J. Biochem. 57:1111-1121; Krude, H. et al. (1998) Nat. Genet. 19:155-157; Online Mendelian Inheritance in Man (OMIM) 176830).

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with membrane proteins for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs). Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin, which function as hormones to regulate cellular functions other than proliferation.

Growth and differentiation factors play critical roles in neoplastic transformation of cells *in vitro* and in tumor progression *in vivo*. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as interferon are cytotoxic to tumor cells both *in vivo* and *in vitro*. Moreover, some growth factors and growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

The Slit protein, first identified in Drosophila, is critical in central nervous system midline

formation and potentially in nervous tissue histogenesis and axonal pathfinding. Itoh et al. (1998; Brain Res. Mol. Brain Res. 62:175-186) have identified mammalian homologues of the slit gene (human Slit-1, Slit-2, Slit-3 and rat Slit-1). The encoded proteins are putative secreted proteins containing EGF-like motifs and leucine-rich repeats, both of which are conserved protein-protein interaction domains. Slit-1, -2, and -3 mRNAs are expressed in the brain, spinal cord, and thyroid, respectively (Itoh et al., *supra*). The Slit family of proteins are indicated to be functional ligands of glypican-1 in nervous tissue and it is suggested that their interactions may be critical in certain stages during central nervous system histogenesis (Liang, Y. et al. (1999) J. Biol. Chem. 274:17885-17892).

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

NP/VMs are involved in numerous neurological and cardiovascular disorders. For example, neuropeptide Y is involved in hypertension, congestive heart failure, affective disorders, and appetite regulation. Somatostatin inhibits secretion of growth hormone and prolactin in the anterior pituitary, as well as inhibiting secretion in intestine, pancreatic acinar cells, and pancreatic beta-cells. A reduction in somatostatin levels has been reported in Alzheimer's disease and Parkinson's disease. Vasopressin acts in the kidney to increase water and sodium absorption, and in higher concentrations stimulates contraction of vascular smooth muscle, platelet activation, and glycogen breakdown in the liver. Vasopressin and its analogues are used clinically to treat diabetes insipidus. Endothelin and angiotensin are involved in hypertension, and drugs, such as captopril, which reduce plasma levels of angiotensin, are used to reduce blood pressure (Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 194; 252; 284; 55; 111).

Neuropeptides have also been shown to have roles in nociception (pain). Vasoactive intestinal peptide appears to play an important role in chronic neuropathic pain. Nociceptin, an endogenous ligand for for the opioid receptor-like 1 receptor, is thought to have a predominantly anti-nociceptive effect, and has been shown to have analgesic properties in different animal models of tonic or chronic pain (Dickinson, T. and S.M. Fleetwood-Walker (1998) Trends Pharmacol. Sci. 19:346-348).

Other proteins that contain signal peptides include secreted proteins with enzymatic activity. Such activity includes, for example, oxidoreductase/dehydrogenase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, or ligase activity. For example, matrix metalloproteinases are secreted hydrolytic enzymes that degrade the extracellular matrix and thus play an important role in tumor metastasis, tissue morphogenesis, and arthritis (Reponen, P. et al. (1995) Dev. Dyn. 202:388-396; Firestein, G.S. (1992) Curr. Opin. Rheumatol. 4:348-354; Ray, J.M. and W.G. Stetler-Stevenson (1994) Eur. Respir. J. 7:2062-2072; and Mignatti, P. and D.B. Rifkin (1993) Physiol. Rev. 73:161-195). Additional examples are the acetyl-CoA synthetases which activate acetate for use in lipid synthesis or energy generation (Luong, A. et al. (2000) J. Biol. Chem. 275:26458-26466). The result of acetyl-CoA synthetase activity is the formation of acetyl-CoA from acetate and CoA. Acetyl-CoA sythetases share a region of sequence similarity identified as the AMP-binding domain signature. Acetyl-CoA synthetase has been shown to be associated with hypertension (Toh, H. (1991) Protein Seq. Data Anal. 4:111-117; and Iwai, N. et al. (1994) Hypertension 23:375-380).

A number of isomerases catalyze steps in protein folding, phototransduction, and various anabolic and catabolic pathways. One class of isomerases is known as peptidyl-prolyl *cis-trans* isomerases (PPIases). PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. Two families of PPIases are the FK506 binding proteins (FKBPs), and cyclophilins (CyPs). FKBPs bind the potent immunosuppressants FK506 and rapamycin, thereby inhibiting signaling pathways in T-cells. Specifically, the PPIase activity of FKBPs is inhibited by binding of FK506 or rapamycin. There are five members of the FKBP family which are named according to their calculated molecular masses (FKBP12, FKBP13, FKBP25, FKBP52, and FKBP65), and localized to different regions of the cell where they associate with different protein complexes (Coss, M. et al. (1995) J. Biol. Chem. 270:29336-29341; Schreiber, S.L. (1991) Science 251:283-287).

The peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. CyP isomerase activity is associated with protein folding and protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsc70 complex that binds steroid receptors. The mammalian CypA has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyPs in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214;

Hunter, T. (1998) Cell 92:141-143; and Leverson, J.D. and S.A. Ness, (1998) Mol. Cell. 1:203-211).

Gamma-carboxyglutamic acid (Gla) proteins rich in proline (PRGPs) are members of a family of vitamin K-dependent single-pass integral membrane proteins. These proteins are characterized by an extracellular amino terminal domain of approximately 45 amino acids rich in Gla. The intracellular carboxyl terminal region contains one or two copies of the sequence PPXY, a motif present in a variety of proteins involved in such diverse cellular functions as signal transduction, cell cycle progression, and protein turnover (Kulman, J.D. et al. (2001) Proc. Natl. Acad. Sci. USA 98:1370-1375). The process of post-translational modification of glutamic residues to form Gla is Vitamin K-dependent carboxylation. Proteins which contain Gla include plasma proteins involved in blood coagulation. These proteins are prothrombin, proteins C, S, and Z, and coagulation factors VII, IX, and X. Osteocalcin (bone-Gla protein, BGP) and matrix Gla-protein (MGP) also contain Gla (Friedman, P.A. and C.T. Przysiecki (1987) Int. J. Biochem. 19:1-80; Vermeer, C. (1990) Biochem. J. 266:625-636).

Immunoglobulins

Antigen recognition molecules are key players in the sophisticated and complex immune systems which all vertebrates have developed to provide protection from viral, bacterial, fungal, and parasitic infections. A key feature of the immune system is its ability to distinguish foreign molecules, or antigens, from "self" molecules. This ability is mediated primarily by secreted and transmembrane proteins expressed by leukocytes (white blood cells) such as lymphocytes, granulocytes, and monocytes. Most of these proteins belong to the immunoglobulin (Ig) superfamily, members of which contain one or more repeats of a conserved structural domain. This Ig domain is comprised of antiparallel β sheets joined by a disulfide bond in an arrangement called the Ig fold. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins and immune cell-specific surface markers such as the "cluster of differentiation" or CD antigens, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of β-sheets. Conserved cysteine residues form an intrachain disulfide-bonded loop, 55-75 amino acid residues in length, which connects the two layers of β-sheets. Each β-sheet has three or four anti-parallel β-strands of 5-10 amino acid residues. Hydrophobic and

hydrophilic interactions of amino acid residues within the β-strands stabilize the Ig fold (hydrophobic on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of β-strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycophosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp. 142-145.)

Antibodies

MHC proteins are cell surface markers that bind to and present foreign antigens to T cells. MHC molecules are classified as either class I or class II. Class I MHC molecules (MHC I) are expressed on the surface of almost all cells and are involved in the presentation of antigen to cytotoxic T cells. For example, a cell infected with virus will degrade intracellular viral proteins and express the protein fragments bound to MHC I molecules on the cell surface. The MHC I/antigen complex is recognized by cytotoxic T-cells which destroy the infected cell and the virus within. Class II MHC molecules are expressed primarily on specialized antigen-presenting cells of the

immune system, such as B-cells and macrophages. These cells ingest foreign proteins from the extracellular fluid and express MHC II/antigen complex on the cell surface. This complex activates helper T-cells, which then secrete cytokines and other factors that stimulate the immune response. MHC molecules also play an important role in organ rejection following transplantation. Rejection occurs when the recipient's T-cells respond to foreign MHC molecules on the transplanted organ in the same way as to self MHC molecules bound to foreign antigen. (Reviewed in Alberts et al., *supra*, pp. 1229-1246.)

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the α, δ, ϵ, γ, and μ H-chain types. There are two types of L-chains, κ and λ, either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts et al. *supra*, pp. 1206-1213; 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-

specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carrry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catcholamines in the central nervous system, and reduce inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex

hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Medroxyprogesterone (MAH), also known as 6α-methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

Mifepristone, also known as RU-486, is an antiprogesterone drug that blocks receptors of progesterone. It counteracts the effects of progesterone, which is needed to sustain pregnancy. Mifepristone induces spontaneous abortion when administered in early pregnancy followed by treatment with the prostaglandin, misoprostol. Further, studies show that mifepristone at a substantially lower dose can be highly effective as a postcoital contraceptive when administered within five days after unprotected intercourse, thus providing women with a "morning-after pill" in case of contraceptive failure or sexual assault. Mifepristone also has potential uses in the treatment of breast and ovarian cancers in cases in which tumors are progesterone-dependent. It interferes with steroid-dependent growth of brain meningiomas, and may be useful in treatment of endometriosis where it blocks the estrogen-dependent growth of endometrial tissues. It may also be useful in treatment of uterine fibroid tumors and Cushing's Syndrome. Mifepristone binds to glucocorticoid

receptors and interferes with cortisol binding. Mifepristone also may act as an anti-glucocorticoid and be effective for treating conditions where cortisol levels are elevated such as AIDS, anorexia nervosa, ulcers, diabetes, Parkinson's disease, multiple sclerosis, and Alzheimer's disease.

Danazol is a synthetic steroid derived from ethinyl testosterone. Danazol indirectly reduces estrogen production by lowering pituitary synthesis of follicle-stimulating hormone and LH. Danazol also binds to sex hormone receptors in target tissues, thereby exhibiting anabolic, antiestrognic, and weakly androgenic activity. Danazol does not possess any progestogenic activity, and does not suppress normal pituitary release of corticotropin or release of cortisol by the adrenal glands. Danazol is used in the treatment of endometriosis to relieve pain and inhibit endometrial cell growth. It is also used to treat fibrocystic breast disease and hereditary angioedema.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A_2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of

the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of β-adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

Histological and molecular evaluation of breast tumors reveals that the development of breast cancer evolves through a multi-step process whereby pre-malignant mammary epithelial cells undergo a relatively defined sequence of events leading to tumor formation. An early event in tumor development is ductal hyperplasia. Cells undergoing rapid neoplastic growth gradually progress to invasive carcinoma and become metastatic to the lung, bone, and potentially other organs. Several variables that may influence the process of tumor progression and malignant transformation include genetic factors, environmental factors, growth factors, and hormones. Based on the complexity of this process, it is critical to study a population of human mammary epithelial cells undergoing the process of malignant transformation, and to associate specific stages of progression with phenotypic and molecular characteristics. We have compared primary breast epithelial cells (HMECs) to breast carcinoma lines at various stages of tumor progression.

HMEC is a primary breast epithelial cell line isolated from a normal donor.

MCF-10A is a breast mammary gland (luminal ductal characteristics) cell line that was isolated from a

36-year-old woman with fibrocystic breast disease. MCF-10A expresses cytoplasmic keratins, epithelial sialomucins, and milkfat globule antigens. This cell lines exhibits three-dimensional growth in

collagen and forms domes in confluent culture.

MCF7 is a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female.

MCF7 has retained characteristics of the mammary epithelium such as the ability to process estradiol via cytoplasmic estrogen receptors and the capacity to form domes in culture.

T-47D is a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast.

Sk-BR-3 is a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female. It forms poorly differentiated adenocarcinoma when injected into nude mice.

BT-20 is a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female.

MDA-mb-231 is a breast tumor cell line isolated from the pleural effusion of a 51-year old female. It forms poorly differentiated adenocarcinoma in nude mice and ALS treated BALB/c mice. It also expresses the Wnt3 oncogene, EGF, and TGF-α .

MDA-mb-435S is a spindle shaped strain that evolved from the parent line (435) as isolated in 1976 by

R. Cailleau from the pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast.

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy, treatment

involves the use of 3–4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30–40% of patients with non-metastatic disease relapse after therapy. Currently, prognostic factor exists that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome.

The most important function of adipose tissue is its ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of excess energy use, and its primary purpose is mobilization during energy deprivation. Understanding how the various molecules regulate adiposity and energy balance in physiological and pathophysiological situations may lead to the development of novel therapeutics for human obesity. Adipose tissue is also one of the important target tissues for insulin. Adipogenesis and insulin resistance in type II diabetes are linked and present intriguing relations. Most patients with type II diabetes are obese and obesity in turn causes insulin resistance. Thiazolidinedione (TZD), a family of drugs of peroxisome proliferation-activated receptor gamma (PPAR-γ) agonists, are a new class of antidiabetic agents that improve insulin sensitivity and reduce plasma glucose and blood pressure in subjects with type II

diabetes. TZD is also able to induce preadipocytes to differentiate into mature fat cells. The majority of research in adipocyte biology to date has been done using transformed mouse preadipocyte cell lines. It has been demonstrated that the culture condition, which stimulates mouse preadipocyte differentiation is different from that for inducing human primary preadipocyte differentiation. In addition, primary cells are diploid and may therefore reflect the *in vivo* context better than aneuploid cell lines.

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer.

For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors and include, simultaneously, widespread genomic hypomethylation and localized areas of increased methylation. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede cancer, and this increase continues during the progression of colonic neoplasms (Wafik, S. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3470-3474). Increased DNA methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal methylation of CpG islands may prevent differentiation in immortalized cells (Antequera, F. et al. (1990) Cell 62:503-514).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. FAP is characterized by the early development of multiple colorectal adenomas that progress to cancer at a mean age of 44 years. The APC gene is a part of the APC-ß-catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. A series of other genetic changes follow activation of the APC-ß-catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation of the K-Ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/ DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC-ß-catenin-Tcf pathway is believed to be central to the

majority of colon cancers in the general population.

Hereditary nonpolyposis Colorectal Cancer (HNPCC) is another inherited autosomal dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach and biliary system. HNPCC results from the mutation of one or more genes in the DNA mis-match repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGFß receptor II gene which controls cell growth. Because of the potential for irreparable damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is characterized by loss of the p53 gene which may occur early, appearing even in histologically normal tissue. The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an intermediate polyp state suggests a high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa, J-P. J. et al. (1994) Nature Genetics 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease, particularly the downregulation or deletion of genes, that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain or lung. About 80% of these tumors remain responsive to androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is currently no known treatment for this condition.

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin J et al. (1999) Cancer Res. 59:2891-2897; Putz T et al. (1999) Cancer Res 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proven useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Leukemias can be classified into four major categories, and all involve malignant transformation of pluripotent stem cells. Acute leukemias, both lymphoblastic (ALL) and myeloid (AML) types, are characterized by the presence of immature cells in the blood. Chronic leukemias, both lymphocytic (CLL) and myelocytic (CML), are associated with mature, differentiated cells, but proportions of each cell type are abnormal. For example, CLL patients usually have clonal expansion of B cell lymphocytes. CML patients often have granulocytes of all stages of maturity present in blood, bone marrow, and other organs. Monoclonal antibodies specific for B- and T-cells are helpful diagnostic tools, in addition to histological analysis. Disease progresses as normal hematopoietic bone marrow is displaced by malignant cells. Cause has been determined to be genetic in some cases, and chemical or radiation-induced in others.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, secreted proteins, referred to collectively as "SECP" and individually as "SECP-1," "SECP-2," "SECP-3," "SECP-4," "SECP-5," "SECP-6," "SECP-7," "SECP-8," "SECP-9," "SECP-10," "SECP-11," "SECP-12," "SECP-13," "SECP-14," "SECP-15," "SECP-16," "SECP-17," "SECP-18," "SECP-19," "SECP-20," "SECP-21," "SECP-22," "SECP-23," "SECP-24," "SECP-25," "SECP-26," "SECP-27," "SECP-28," "SECP-29," "SECP-30," "SECP-31," "SECP-32," "SECP-33," "SECP-34," "SECP-35," "SECP-36," "SECP-37," "SECP-38," "SECP-39," "SECP-40," "SECP-41," "SECP-42," "SECP-43," "SECP-44," "SECP-45," "SECP-46," "SECP-47," "SECP-48," "SECP-49," "SECP-50," "SECP-51," "SECP-52," "SECP-53," "SECP-54," "SECP-55," "SECP-56," "SECP-57," "SECP-58," "SECP-59," "SECP-60," "SECP-61," "SECP-62," "SECP-63," "SECP-64," "SECP-65," "SECP-66," "SECP-67," "SECP-68," "SECP-69," "SECP-70," "SECP-71," "SECP-72," "SECP-73," "SECP-74," "SECP-75," "SECP-76," "SECP-77," "SECP-78," "SECP-79," and "SECP-80," and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical

conditions. Embodiments also provide methods for utilizing the purified secreted proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified secreted proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-80.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-80. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:81-160.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group

consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe

specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

selected from the group consisting of SEQ ID NO:1-80. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, c) a biologically active

fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of

hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide embodiments, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"SECP" refers to the amino acid sequences of substantially purified SECP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of SECP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

An "allelic variant" is an alternative form of the gene encoding SECP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding SECP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as SECP or a polypeptide with at least one functional characteristic of SECP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SECP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding SECP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent SECP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SECP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar

hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of SECP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind SECP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries.

Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic SECP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given

polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding SECP or fragments of SECP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,

(b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of SECP or a polynucleotide encoding SECP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:81-160 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:81-160, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:81-160 can be

employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:81-160 from related polynucleotides. The precise length of a fragment of SEQ ID NO:81-160 and the region of SEQ ID NO:81-160 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-80 is encoded by a fragment of SEQ ID NO:81-160. A fragment of SEQ ID NO:1-80 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-80. For example, a fragment of SEQ ID NO:1-80 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-80. The precise length of a fragment of SEQ ID NO:1-80 and the region of SEQ ID NO:1-80 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the

percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific

hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of SECP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of SECP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of SECP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of SECP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an SECP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will

vary by cell type depending on the enzymatic milieu of SECP.

"Probe" refers to nucleic acids encoding SECP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999) Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific

needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing SECP, nucleic acids encoding SECP, or fragments thereof may comprise a bodily fluid; an extract from a

cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with

a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human secreted proteins (SECP), the polynucleotides encoding SECP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to polypeptide and polynucleotide embodiments. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptides shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these

properties establish that the claimed polypeptides are secreted proteins. For example, SEQ ID NO:16 is 71% identical, from residue M1 to residue D238, to human C1q-related factor (GenBank ID g3747097) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.9e-91, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a C1q domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:16 is a C1q-related complement factor. In an alternative example, SEQ ID NO:28 is 41% identical, from residue M1 to residue L120, to Rattus norvegicus Ly6C antigen (GenBank ID g205250) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.0e-18, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:28 also contains a signal peptide and a u-Par/Ly-6 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis and from BLAST analysis of the DOMO database provide further corroborative evidence that SEQ ID NO:28 is a secreted antigen. In an alternative eample, SEQ ID NO:29 is 78% identical, from residue G66 to residue D129, to human PAP (pancreatitis associated protein) homologous protein (GenBank ID g285971) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.5e-58, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Pancreatitis associated protein I is a secretory stress protein first characterized in pancreas during pancreatitis but also expressed in several tissues including hepatic, gastric, and colon cancer. Its concentration in serum can be significant. Exogenous pancreatitis associated protein I can modify the adhesion and motility of normal and transformed melanocytes, suggesting a potential interaction with melanoma invasivity (Valery C et al (2001) J Invest Dermatol 116(3):426-433.) SEQ ID NO:29 also contains a lectin C-type domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, PROFILESCAN, BLIMPS, and further BLAST analyses provide corroborative evidence that SEQ ID NO:29 is a PAP homologous protein. In an alternative example, SEQ ID NO:45 is 78% identical, from residue G66 to residue D129, and 87% identical, from residue M1 to residue D65, to human pancreatitis-associated protein (GenBank ID g482909) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.5e-58, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:45 also contains a lectin C-type

domain and a signal peptide as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, PROFILESCAN, and MOTIFS analyses and BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:45 is a secreted lectin-related protein. In an alternative example, SEQ ID NO:58 is 98% identical, from residue D28 to residue L115 and 100% identical, from residue M1 to residue C27, to macaque epididymal secretory protein, ESP14.6 (GenBank ID g794071) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.6e-56, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:58 also contains a E1 family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) SEQ ID NO:1-15, SEQ ID NO:17-27, SEQ ID NO:30-44, SEQ ID NO:46-57, and SEQ ID NO:59-80 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-80 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:81-160 or that distinguish between SEQ ID NO:81-160 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon

stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to

assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide embodiments, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses SECP variants. A preferred SECP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the SECP amino acid sequence, and which contains at least one functional or structural characteristic of SECP.

Various embodiments also encompass polynucleotides which encode SECP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:81-160, which encodes SECP. The polynucleotide sequences of SEQ ID NO:81-160, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding SECP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding SECP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:81-160 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-160. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of SECP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding SECP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding SECP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding SECP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding SECP. For example, a polynucleotide comprising a sequence of SEQ ID NO:153, a polynucleotide comprising a sequence of SEQ ID NO:154, a polynucleotide comprising a sequence of SEQ ID NO:155, a polynucleotide comprising a sequence of SEQ ID NO:156, a polynucleotide comprising a sequence of SEQ IDNO:157, a polynucleotide comprising a sequence of SEQ ID NO:158, and a polynucleotide comprising a sequence of SEQ ID NO:159 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:111 and a polynucleotide comprising a sequence of SEQ ID NO:116, are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:160 and a polynucleotide comprising a sequence of SEQ ID NO:152 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of SECP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding SECP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring SECP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode SECP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring SECP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding SECP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SECP and its derivatives without altering the encoded amino acid sequences include the

production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode SECP and SECP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding SECP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:81-160 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding SECP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known

sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode SECP may be cloned in recombinant DNA molecules that direct expression of SECP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express SECP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter SECP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed

mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of SECP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding SECP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, SECP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of SECP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (Creighton, *supra*, pp. 28-53).

In order to express a biologically active SECP, the polynucleotides encoding SECP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains

the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding SECP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding SECP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding SECP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding SECP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding SECP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci.

USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding SECP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding SECP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding SECP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of SECP are needed, e.g. for the production of antibodies, vectors which direct high level expression of SECP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of SECP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of SECP. Transcription of polynucleotides encoding SECP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding SECP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to

obtain infective virus which expresses SECP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of SECP in cell lines is preferred. For example, polynucleotides encoding SECP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding SECP is inserted within a marker gene sequence, transformed cells containing

polynucleotides encoding SECP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SECP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding SECP and that express SECP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of SECP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SECP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SECP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding SECP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding SECP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors

containing polynucleotides which encode SECP may be designed to contain signal sequences which direct secretion of SECP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding SECP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric SECP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of SECP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the SECP encoding sequence and the heterologous protein sequence, so that SECP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled SECP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

SECP, fragments of SECP, or variants of SECP may be used to screen for compounds that specifically bind to SECP. One or more test compounds may be screened for specific binding to SECP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to SECP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of SECP can be used to screen for binding of test compounds, such as antibodies, to SECP, a variant of SECP, or a combination of SECP and/or one or more variants SECP. In an embodiment, a variant of SECP can be used to screen for compounds that bind to a variant of SECP, but not to SECP having the exact sequence of a sequence of SEQ ID NO:1-80. SECP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to SECP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to SECP can be closely related to the natural ligand of SECP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor SECP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to SECP can be closely related to the natural receptor to which SECP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for SECP which is capable of propagating a signal, or a decoy receptor for SECP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to SECP, fragments of SECP, or variants of SECP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of SECP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of SECP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential

diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of SECP.

In an embodiment, anticalins can be screened for specific binding to SECP, fragments of SECP, or variants of SECP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit SECP involves producing appropriate cells which express SECP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing SECP or cell membrane fractions which contain SECP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either SECP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with SECP, either in solution or affixed to a solid support, and detecting the binding of SECP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991)

Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

SECP, fragments of SECP, or variants of SECP may be used to screen for compounds that modulate the activity of SECP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for SECP activity, wherein SECP is combined with at least one test compound, and the activity of SECP in the presence of a test compound is compared with the activity of SECP in the absence of the test compound. A change in the activity of SECP in the presence of the test compound is indicative of a compound that modulates the activity of SECP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising SECP under conditions suitable for SECP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of SECP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding SECP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding SECP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding SECP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding SECP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress SECP, e.g., by secreting SECP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of SECP and secreted proteins. In addition, examples of tissues expressing SECP can be found in Table 6 and can also be found in Example XI. Therefore, SECP appears to play a role in cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders. In the treatment of disorders associated with increased SECP expression or activity, it is desirable to decrease the expression or activity of SECP. In the treatment of disorders associated with decreased SECP expression or activity, it is desirable to increase the expression or activity of SECP.

Therefore, in one embodiment, SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or

pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary

mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified SECP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of SECP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those listed above.

In a further embodiment, an antagonist of SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders described above. In one aspect, an antibody which specifically binds SECP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express SECP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of SECP may be produced using methods which are generally known in the art. In particular, purified SECP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SECP. Antibodies to SECP may also be generated using methods that are well known in the art. Such antibodies may include, but are not

limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with SECP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SECP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of SECP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to SECP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SECP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for SECP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between SECP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SECP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for SECP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of SECP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple SECP epitopes, represents the average affinity, or avidity, of the antibodies for SECP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular SECP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the SECP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of SECP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml,

preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of SECP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding SECP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding SECP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding SECP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding SECP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D.

(1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in SECP expression or regulation causes disease, the expression of SECP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in SECP are treated by constructing mammalian expression vectors encoding SECP and introducing these vectors by mechanical means into SECP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of SECP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). SECP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding SECP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to SECP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding SECP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding SECP to cells which have one or more genetic abnormalities with respect to the expression of SECP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding SECP to target cells which have one or more genetic abnormalities with respect to the expression of SECP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing SECP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding SECP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for SECP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of SECP-coding RNAs and the synthesis of high levels of SECP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of SECP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful

because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding SECP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding SECP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding SECP. Compounds

which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased SECP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding SECP may be therapeutically useful, and in the treatment of disorders associated with decreased SECP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding SECP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding SECP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding SECP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding SECP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of SECP, antibodies to SECP, and mimetics, agonists, antagonists, or inhibitors of SECP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising SECP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, SECP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to

transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SECP or fragments thereof, antibodies of SECP, and agonists, antagonists or inhibitors of SECP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind SECP may be used for the diagnosis of disorders characterized by expression of SECP, or in assays to monitor patients being

treated with SECP or agonists, antagonists, or inhibitors of SECP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for SECP include methods which utilize the antibody and a label to detect SECP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring SECP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of SECP expression. Normal or standard values for SECP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to SECP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of SECP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding SECP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of SECP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of SECP, and to monitor regulation of SECP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding SECP or closely related molecules may be used to identify nucleic acid sequences which encode SECP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding SECP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the SECP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:81-160 or from genomic sequences including promoters, enhancers, and introns of the SECP gene.

Means for producing specific hybridization probes for polynucleotides encoding SECP include the cloning of polynucleotides encoding SECP or SECP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding SECP may be used for the diagnosis of disorders associated with expression of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal

disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. Polynucleotides encoding SECP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered SECP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding SECP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding SECP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding SECP in the sample indicates the presence of the associated disorder. Such assays may also be used

to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SECP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding SECP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SECP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding SECP, or a fragment of a polynucleotide complementary to the polynucleotide encoding SECP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding SECP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding SECP are used to amplify DNA using the polymerase chain reaction (PCR).

The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of SECP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be

used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, SECP, fragments of SECP, or antibodies specific for SECP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or

signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed

by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for SECP to quantify the levels of SECP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding SECP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding SECP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further

investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, SECP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SECP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with SECP, or fragments thereof, and washed. Bound SECP is then detected by methods well known in the art. Purified SECP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding SECP specifically compete with a test compound for binding SECP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with SECP.

In additional embodiments, the nucleotide sequences which encode SECP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No.60/313,249, including U.S. Ser. No.60/314,752, U.S. Ser. No.60/317,824, U.S. Ser. No.60/ 317,818, U.S. Ser. No.60/ 324,586, U.S. Ser. No.60/ 362,439, U.S. Ser. No.60/ 357,002, U.S. Ser. No.60/ 343,980, U.S. Ser. No.60/ 334,229, U.S. Ser. No.60/ 366,041, U.S. Ser. No.60/ 376,988, and U.S. Ser. No.60/ 324,040 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96

plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus

primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio Inc., Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:81-160. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative secreted proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of

organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode secreted proteins, the encoded polypeptides were analyzed by querying against PFAM models for secreted proteins. Potential secreted proteins were also identified by homology to Incyte cDNA sequences that had been annotated as secreted proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent

type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of SECP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:81-160 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:81-160 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site

(http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding SECP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ

cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding SECP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of SECP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in SECP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:81-160 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated

algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:81-160 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech,

Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibited at least about a two-fold change in expression, a signal-to-

background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed.

Expression

The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. For example, the human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of results obtained using the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein iii) conversion of ammonia to urea and glutamine; iv) ability to metabolize aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is well established as an in vitro model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416). SEQ ID NO:96 showed differential expression in C3A cells treated with a variety of steroids including beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, and progesterone, versus untreated C3A cells, as determined by microarray analysis. Specifically, the expression of SEQ ID NO:96 was increased at least 2-fold by treatment of cells with: 1-100 microM medroxyprogesterone for 1-6 hours, 1-100 microM budesonide for 1-6 hours, 1-100 microM progesterone for 1 hour, and 1-100 microM betamethasone Therefore, SEQ ID NO:96 is useful for the diagnosis and monitoring of liver, endocrine, and reproductive diseases and in the diagnosis of and as a therapeutic target for inflammatory diseases and humoral immune response.

In an alternative example, the gene expression profile of nonmalignant primary mammary epithelial cells (HMECs) was compared to that of various breast carcinoma lines at different stages of tumor progression. The breast carcinoma lines studied were BT-20, MCF7, MDA-mb-435S, Sk-BR-3, and T-47D. SEQ ID NO:115 was found to be downregulated by at least two-fold in MCF7, Sk-BR-3, and T-47D. Therefore SEQ ID NO:115, encoding SEQ ID NO:35 can be used in assays to detect breast cancer.

In an alternative example, SEQ ID NO:115 was downregulated by at least two-fold in osteosarcoma tissues when compared to its expression in normal osteoblast primary culture cells, the NHO_{st}5488 cells, in four out of the seven donors studied. Therefore, SEQ ID NO:115, encoding SEQ ID NO:35 can be used in assays to detect osteosarcoma.

In an alternative example, human preadipocytes were treated with human insulin and PPAR-γ agonist for 3 days and subsequently were switched to medium containing insulin for 24 hours, 48 hours, 4 days, 1.1 week, and 2.1 weeks before the cells were collected for analysis. Differentiated

adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. SEQ ID NO:115 was downregulated by at least two-fold in the differentiated adipocytes after a minimum of 48 hours in the medium containing insulin and remained so for a maximum of 1.1 week. Therefore, through learning the gene expression profile during adipogenesis in humans, it will be possible to understand the fundamental mechanism of adiposity regulation. Furthermore, by comparing the gene expression profiles of adipogenesis in normal weight and donor with obesity it will be possible to identify crucial genes, which might be potential drug targets for obesity and type II diabetes. SEQ ID NO:115, encoding SEQ ID NO:35 can be used in the above assays.

In an alternative example, SEQ ID NO:125 showed differential expression in colon tissue from patients with colon cancer compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. The expression of SECP-45 was increased at least two-fold in cancerous colon tissue. SEQ ID NO:125 also showed differential expression in prostate LNCaP carcinoma cells compared to prostate PrEC epithelial cells as determined by microarray analysis. The LNCaP cell line was isolated from a lymph node biopsy of a 50-year old male with metastatic prostate carcinoma. The expression of SECP-45 was decreased at least two-fold in prostate LNCaP carcinoma cells compared to prostate PrEC epithelial cells. In an alternative example, SEQ ID NO:125 showed differential expression associated with immune and inflammatory responses as determined by microarray analysis. The expression of SEQ ID NO:125 was increased by at least two-fold in peripheral blood mononuclear cells (PBMCs; 12% B lymphocytes, 40% T lymphocytes, 20% NK cells, 25% monocytes, and 3% various cells that include dendritic and progenitor cells) treated with interleukin-1β (IL-1β), interleukin-6 (IL-6), or tumor necrosis factor-α (TNF-α) compared to untreated PBMCs. IL-1β is a cytokine that plays roles in acute inflammatory responses, fever induction, metabolic regulation, and bone remodeling. IL-1β induces its own production in monocytes and also induces production of adhesion molecules and chemokines in endothelial cells and interferon-γ in NK cells. IL-6 plays roles in host defense, immune responses, and hematopoiesis. TNF-α is a pleiotropic cytokine involved in immune regulation and inflammatory responses. SEQ ID NO:125 also showed at least 2-fold decreased expression in human T cell leukemia Jurkat cells treated with a combination of the protein kinase C activator, phorbol myristate acetate (PMA), and the calcium ionophore, ionomycin, compared to untreated Jurkat cells as determined by microarray analysis. Treatment of T cells with PMA and ionomycin mimics the signaling events elicited during T cell activation. In addition, SEQ ID NO:125 showed at least two-fold decreased expression in THP-1 promonocyte cells stimulated with PMA and ionomycin. THP-1 is a promonocyte cell line isolated from the peripheral blood of a 1-year-old male with acute monocytic leukemia. THP-1 cells acquire monocytic characteristics in response to stimulation with

PMA. Therefore, SEQ ID NO:125 is useful in disease staging and diagnostic assays for cell proliferative disorders, including breast cancer, colon cancer, and prostate cancer, and autoimmune/inflammatory disorders.

In an alternative example, SEQ ID NO:128 showed differential expression in brain cingulate from a patient with Alzheimer's disease compared to matched microscopically normal tissue from the same donor as determined by microarray analysis. The expression of SECP-48 was increased at least two-fold in cingulate tissue with Alzheimer's disease. Therefore, SEQ ID NO:128 is useful in disease staging and diagnostic assays for neurological disorders, including Alzheimer's disease.

In an alternative example, human LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a male donor with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens and androgen receptors, and produce prostatic acid phosphatase.. PrEC is a primary prostate epithelial cell line isolated from a normal donor. In LNCaP cells, one of three metastatic prostate carcinoma cell lines tested, SEQ ID NO:138 was downregulated at least two-fold when compared with PrEC cells.

In an alternative example, Jurkat is an acute T cell leukemia cell line that grows actively in the absence of external stimuli. Jurkat has been extensively used to study signaling in human T cells. PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In T cells, the combination of PMA and ionomycin mimics the type of secondary signaling events elicited during optimal B cell activation. SEQ ID NO:149 was downregulated at least two-fold in the Jurkat T-cell leukemia cell line that had been stimulated for one hour with 1μM PMA (phorbol 12-myristate 13-acetate) and with ionomycin.concentrations varying between 50 ng/ml and 1 μg/ml when compared to untreated Jurkat cells in the absence of stimuli.

In an alternative example, THP-1 is a promonocyte cell line that was isolated from the peripheral blood of a 1-year-old male with acute monocytic leukemia. Upon stimulation with PMA, THP-1 differentiates into a macrophage-like cell that displays many characteristics of peripheral human macrophages. THP-1 cells have been extensively used in the study of signaling in human monocytes and the identification of new factors produced by human monocytes. SEQ ID NO:150 was downregulated at least two-fold in THP-1 cells that had been stimulated for four or more hours with 0.1 μM PMA and then further stimulated with 1 μg/ml ionomycin when compared to untreated THP-1 cells in the absence of stimuli. Also, SEQ ID NO:150 was upregulated at least two-fold in osteosarcoma tissue from two donors with chondroblastic osteosarcoma of the femur when compared with a normal osteoblast cell line.

In an alternative example, a pure human mammary epithelial cell (HMEC) population was compared to breast carcinoma lines at various stages of tumor progression. SEQ ID NO:156 was found to be downregulated at least two fold in BT-20, BT-474, BT-483, Hs578T, MCF7, MDA-MB-468. Therefore SEQ ID NO:156 can be used in assays to detect breast cancer.

In an alternative example, the aim was to identify genes differentially regulated during the process of tumor progression. To this end, the gene expression profiles of primary prostate epithelial cells and prostate carcinomas that are representative of the different stages of tumor progression were compared. SEQ ID NO:156 was found to be downregulated at least two fold in DU 145, LNCaP, and PC-3. Therefore, SEQ ID NO:156 can be used in assays to detect prostrate cancer.

In an alternative example, SEQ ID NO:157 showed differential expression associated with breast cancer, as determined by microarray analysis. Breast carcinoma cell lines at various stages of tumor progression were compared to primary human breast epithelial cells. The breast carcinoma cell lines include MCF7, a breast adenocarcinoma cell line derived from the pleural effusion of a 69-year-old female; T-47D, a breast carcinoma cell line derived from a pleural effusion from a 54-year-old female with an infiltrating ductal carcinoma of the breast; Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female; BT-20, a breast adenocarcinoma isolated in vitro from cells emigrating out of thin slices of a tumor mass isolated from a 74-year-old female; MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, which forms poorly differentiated adenocarcinoma in nude mice and expresses the Wnt3 oncogene, EGF and TGF-α; and MDA-mb-435S, a spindle shaped strain that evolved from a cell line isolated from the pleural effusion of a 31 year old female with metastatic, ductal adenocarcinoma of the breast. The nonmalignant breast epithelial cell line, MCF-10A was isolated from a 36-year-old woman with fibrocystic breast disease. All cell cultures were propagated in a chemically-defined medium, according to the supplier's recommendations and grown to 70-80% confluence prior to RNA isolation. The microarray experiments showed that expression of SEQ ID NO:157 was up-regulated by at least two-fold in two of six cell lines examined as compared to the nonmalignant breast epithelial cell line, MCF-10A.

In another experiment designed to investigate the process of tumor progression and malignant transformation in breast tumors, a comparison was made against the primary mammary epithelial cell line HMEC, derived from normal human mammary tissue (Clonetics, San Diego, CA) and the breast tumor cell lines described above. The microarray experiments indicated that expression of SEQ ID NO:157 was decreased by at least two fold in five breast tumor cells lines when compared to HMEC cells. Therefore, SEQ ID NO:157 is useful in diagnostic and disease staging assays for breast cancer and as a potential biological marker and therapeutic agent in the treatment of breast cancer.

In an alternative example, SEQ ID NO:157 also showed differential expression in prostate cancer, as determined by microarray analysis. Prostate carcinoma cell lines at various stages of tumor progression were compared to primary prostate epithelial cells. The prostate carcinoma cell lines include: DU145, a prostate carcinoma cell line with no detectable sensitivity to hormones, isolated from a metastatic site in the brain of a 69-year-old male, that does not express prostate specific antigen; LNCaP, a prostate carcinoma cell line that expresses androgen receptors and prostate specific antigen and was isolated from a lymph node of a 50-year-old male with metastatic prostate cancer; and PC3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma. The primary prostate epithelial cells, PrECs, were isolated from a normal donor. All cell cultures were propagated in a chemically-defined medium, according to the supplier's recommendations and grown to 70-80% confluence prior to RNA isolation. The microarray experiments showed that expression of SEQ ID NO:157 was decreased by at least two-fold in all three prostate carcinoma cell lines, as well as in prostate metastatic samples from brain, bone and nodes, as compared to primary prostate epithelial cells. Therefore, SEQ ID NO:157 is useful in diagnostic and staging assays for prostate cancer and as a potential biological marker and therapeutic agent in the treatment of prostate cancer.

In an alternative example, SEQ ID NO:158 showed differential expression in breast carcinoma cell lines versus primary mammary epithelial cells as determined by microarray analysis. The breast carcinoma cell lines include BT20, a breast carcinoma cell line derived in vitro from cells emigrating out of thin slices of a tumor mass isolated from a 74-year-old female; BT474, a breast ductal carcinoma cell line isolated from a solid, invasive ductal carcinoma of the breast from a 60-year-old female; BT483, a breast ductal carcinoma cell line isolated from a papillary invasive ductal tumor from a 23-year-old normal, menstruating, parous female; HS578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma; MCF7, a breast adenocarcinoma cell line derived from the pleural effusion of a 69-year-old female; and MDA-mb-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast. The primary mammary epithelial cell line HMEC was derived from normal human mammary tissue (Clonetics, San Diego, CA). The microarray experiments showed that the expression of SEQ ID NO:158 were decreased by at least four fold in all six breast carcinoma lines (BT20, BT474, BT483, HS578T, MCF7, and MDA-mb-468) relative to cells from the primary mammary epithelial cell line, HMEC. Therefore, SEQ ID NO:158 is useful as a diagnostic marker or as a potential therapeutic target for breast cancer.

In an alternative example, SEQ ID NO:158 also showed differential expression in prostate carcinoma cell lines versus normal prostate epithelial cells as determined by microarray analysis. Three prostate carcinoma cell lines, DU 145, LNCaP, and PC-3 were included in the experiments.

DU 145 was isolated from a metastatic site in the brain of a 69 year old male with widespread metastatic prostate carcinoma. DU 145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weekly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA). LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50 year old male with metastatic prostate carcinoma. LNCaP expresses PSA, produces prostate acid phosphatase, and expresses androgen receptors. PC-3, a prostate adenocarcinoma cell line, was isolated from a metastatic site in the bone of a 62 year old male with grade IV prostate adenocarcinoma. The normal epithelial cell line, PrEC, is a primary prostate epithelial cell line isolated from a normal donor. In one experiment, the expression of cDNAs from the prostate carcinoma cell lines were compared to that of the normal prostate epithelial cells grown under the same conditions (in the absence of growth factors and hormones). This experiment showed that the expression of SEQ ID NO:158 was decreased by at least four fold in both all three prostate carcinoma lines relative to PrECs. In the other experiment, the expression of cDNAs from the prostate carcinoma cell lines grown in optimal conditions (in the presence of growth factors and hormones) were compared to that of the normal prostate epithelial cells grown under restrictive conditions (in the absence of growth factors and hormones). The experiment showed that the expression of SEQ ID NO:158 was also decreased by at least four fold in DU145, LNCaP, and PC-3 prostate carcinoma lines relative to PrECs. Therefore, SEQ ID NO:158 is useful as a diagnostic marker or as a potential therapeutic target for prostate cancers.

XII. Complementary Polynucleotides

Sequences complementary to the SECP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring SECP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of SECP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the SECP-encoding transcript.

XIII. Expression of SECP

Expression and purification of SECP is achieved using bacterial or virus-based expression systems. For expression of SECP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express SECP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of SECP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding SECP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, SECP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from SECP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified SECP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX where applicable.

XIV. Functional Assays

SECP function is assessed by expressing the sequences encoding SECP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser

optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of SECP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding SECP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding SECP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of SECP Specific Antibodies

SECP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the SECP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-SECP activity by, for example, binding the peptide or SECP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring SECP Using Specific Antibodies

Naturally occurring or recombinant SECP is substantially purified by immunoaffinity chromatography using antibodies specific for SECP. An immunoaffinity column is constructed by covalently coupling anti-SECP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing SECP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SECP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/SECP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and SECP is collected.

XVII. Identification of Molecules Which Interact with SECP

SECP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SECP, washed, and any wells with labeled SECP complex are assayed. Data obtained using different concentrations of SECP are used to calculate values for the number, affinity, and association of SECP with the candidate molecules.

Alternatively, molecules interacting with SECP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

SECP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of SECP Activity

An assay for growth stimulating or inhibiting activity of SECP measures the amount of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY). In this assay, varying amounts of SECP are added to quiescent 3T3 cultured cells in the presence of [^{3}H]thymidine, a radioactive DNA precursor. SECP for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold SECP concentration range is indicative of growth modulating activity. One unit of activity per milliliter is defined as the

concentration of SECP producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA .

Alternatively, an assay for SECP activity measures the stimulation or inhibition of neurotransmission in cultured cells. Cultured CHO fibroblasts are exposed to SECP. Following endocytic uptake of SECP, the cells are washed with fresh culture medium, and a whole cell voltage-clamped *Xenopus* myocyte is manipulated into contact with one of the fibroblasts in SECP-free medium. Membrane currents are recorded from the myocyte. Increased or decreased current relative to control values are indicative of neuromodulatory effects of SECP (Morimoto, T. et al. (1995) Neuron 15:689-696).

Alternatively, an assay for SECP activity measures the amount of SECP in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations from fractionated and total cell lysates are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of SECP in secretory organelles relative to SECP in total cell lysate is proportional to the amount of SECP in transit through the secretory pathway.

Alternatively, AMP binding activity is measured by combining SECP with ^{32}P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to SECP activity.

XIX. Demonstration of Immunoglobulin Activity

An assay for SECP activity measures the ability of SECP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction. (Golub, E.S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland, MA, pp. 113-115.) SECP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled SECP. SECP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable SECP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable SECP-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable SECP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the

amount of precipitable SECP-antigen complex is a measure of SECP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for SECP activity measures the expression of SECP on the cell surface. cDNA encoding SECP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of SECP expressed on the cell surface.

Alternatively, an assay for SECP activity measures the amount of cell aggregation induced by overexpression of SECP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding SECP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of SECP activity.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
1417062	1	1417062CD1	81	1417062CB1	90110408CA2, 90110416CA2, 90110432CA2, 90172377CA2, 90172385CA2, 90172453CA2, 90172461CA2, 90172477CA2, 90172485CA2
2007701	2	2007701CD1	82	2007701CB1	
2915695	3	2915695CD1	83	2915695CB1	
2969449	4	2969449CD1	84	2969449CB1	
2994102	5	2994102CD1	85	2994102CB1	90110580CA2, 90110588CA2, 90110596CA2
3410251	6	3410251CD1	86	3410251CB1	90126170CA2, 90126194CA2, 90126262CA2
5330327	7	5330327CD1	87	5330327CB1	
5532048	8	5532048CD1	88	5532048CB1	
56002716	9	56002716CD1	89	56002716CB1	56002716CA2, 90110503CA2, 90110511CA2, 90110535CA2, 90110619CA2
60129797	10	60129797CD1	90	60129797CB1	90109831CA2, 90109847CA2
6246243	11	6246243CD1	91	6246243CB1	90188842CA2, 90188850CA2, 90188874CA2, 90188890CA2, 90188902CA2, 90188982CA2
6804755	12	6804755CD1	92	6804755CB1	6804755CA2, 90125796CA2
6856852	13	6856852CD1	93	6856852CB1	6856852CA2, 90166714CA2, 90166730CA2, 90166738CA2, 90166746CA2, 90166814CA2, 90166822CA2, 90166830CA2, 90166838CA2
7482027	14	7482027CD1	94	7482027CB1	
7493507	15	7493507CD1	95	7493507CB1	
3075994	16	3075994CD1	96	3075994CB1	90164903CA2, 90164911CA2, 90164927CA2, 90165019CA2, 90165027CA2, 90165043CA2
2378119	17	2378119CD1	97	2378119CB1	1483648CA2, 2075676CA2, 2378119CA2, 90166558CA2, 90166582CA2, 90166690CA2, 90166757CA2
2987418	18	2987418CD1	98	2987418CB1	2987418CA2, 5476092CA2, 90166703CA2, 90166803CA2, 90166827CA2
4223862	19	4223862CD1	99	4223862CB1	1597349CA2, 4223862CA2, 90166968CA2, 90166992CA2, 90167185CA2, 90167277CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
6046406	20	6046406CD1	100	6046406CB1	6046406CA2, 90166102CA2, 90166118CA2, 90166134CA2, 90166218CA2, 90166226CA2
6743529	21	6743529CD1	101	6743529CB1	2474053CA2, 6743529CA2, 90166357CA2
7283809	22	7283809CD1	102	7283809CB1	7283809CA2
7637563	23	7637563CD1	103	7637563CB1	7637563CA2
7663814	24	7663814CD1	104	7663814CB1	90166178CA2, 90166286CA2
8001939	25	8001939CD1	105	8001939CB1	90173305CA2, 90173313CA2, 90173337CA2, 90173421CA2, 90173429CA2, 90173437CA2, 90173445CA2, 90188622CA2, 90188654CA2, 90188662CA2, 90188670CA2, 90188678CA2, 90188686CA2, 90188694CA2, 90188778CA2, 90188786CA2, 90188938CA2
8191019	26	8191019CD1	106	8191019CB1	90166811CA2, 90166819CA2, 90166835CA2, 90167102CA2, 90167118CA2, 90167126CA2, 90167134CA2, 90167142CA2, 90167202CA2, 90167210CA2, 90167218CA2, 90167226CA2, 90188610CA2, 90188649CA2,
					90188657CA2, 90188665CA2, 90188673CA2, 90188681CA2, 90188689CA2, 90188749CA2, 90188757CA2, 90188765CA2, 90188773CA2, 90188774CA2, 90188781CA2, 90188789CA2, 90188809CA2, 90188941CA2
919788	27	919788CD1	107	919788CB1	90177394CA2
4758058	28	4758058CD1	108	4758058CB1	
7499835	29	7499835CD1	109	7499835CB1	90108956CA2, 90108988CA2, 90109132CA2, 90109148CA2, 90109156CA2, 90109208CA2, 90109433CA2, 90110058CA2, 90132951CA2, 90132967CA2, 90132975CA2, 90132991CA2, 90133051CA2, 90133059CA2,
					90133067CA2, 90133075CA2, 90133091CA2, 90133161CA2, 90133193CA2, 90133709CA2, 90133725CA2, 90133733CA2, 90133741CA2, 90133817CA2, 90133825CA2, 90133833CA2, 90135904CA2, 90135912CA2, 90135944CA2, 90136020CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
2484647	30	2484647CD1	110	2484647CB1	1726009CA2, 1830692CA2, 6574257CA2, 8612535CA2, 90132931CA2, 90132955CA2, 90132963CA2, 90132971CA2, 90133008CA2, 90133039CA2, 90133055CA2, 90133063CA2, 90133087CA2, 90133103CA2,
					90133123CA2, 90133127CA2, 90133215CA2, 90133219CA2, 90133224CA2, 90133235CA2, 90133251CA2, 90133311CA2
2587034	31	2587034CD1	111	2587034CB1	
2702991	32	2702991CD1	112	2702991CB1	2702991CA2
2744736	33	2744736CD1	113	2744736CB1	2744736CA2, 90133729CA2, 90133745CA2, 90133845CA2
2915475	34	2915475CD1	114	2915475CB1	2915475CA2, 90132915CA2, 90133015CA2, 90133022CA2, 90133023CA2, 90133031CA2, 90133150CA2
3040427	35	3040427CD1	115	3040427CB1	1824963CA2
7499722	36	7499722CD1	116	7499722CB1	
6776909	37	6776909CD1	117	6776909CB1	7985313CA2, 90132902CA2, 90132910CA2, 90132918CA2, 90133002CA2, 90133010CA2, 90133018CA2, 90133034CA2, 90133044CA2, 90133953CA2
7280438	38	7280438CD1	118	7280438CB1	7280438CA2, 8018238CA2, 90133544CA2, 90133644CA2
7499809	39	7499809CD1	119	7499809CB1	
7499921	40	7499921CD1	120	7499921CB1	
2705858	41	2705858CD1	121	2705858CB1	55115172CA2, 90109710CA2, 90109766CA2, 90109774CA2, 90109866CA2, 90109902CA2, 90109909CA2, 90109917CA2, 90109925CA2, 90109933CA2, 90109941CA2, 90109949CA2, 90109957CA2, 90109965CA2, 90109973CA2,
					90109981CA2, 90109989CA2, 90110002CA2, 90110025CA2, 90110041CA2, 90110057CA2, 90110065CA2, 90110073CA2, 90110081CA2, 90110089CA2, 90110450CA2, 90175707CA2, 90175715CA2, 90175739CA2, 90175807CA2, 90175823CA2, 90175831CA2, 90175847CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
3069892	42	3069892CD1	122	3069892CB1	90160650CA2, 90160666CA2, 90160690CA2, 90160758CA2, 90160766CA2, 90160818CA2, 90160826CA2, 90160834CA2, 90160902CA2, 90160910CA2, 90160918CA2, 90160926CA2, 90160934CA2, 90160942CA2, 90187851CA2,
					90187867CA2, 90187891CA2, 90187951CA2, 90187967CA2, 90187991CA2, 90188268CA2, 90188276CA2, 90188284CA2, 90188292CA2, 90188360CA2, 90188368CA2, 90188384CA2, 90188428CA2, 90188451CA2, 90188459CA2,
					90188465CA2, 90188467CA2, 90188475CA2, 90188483CA2, 90188491CA2, 90188551CA2, 90188559CA2, 90188567CA2, 90188575CA2, 90188583CA2, 90188591CA2
3069586	43	3069586CD1	123	3069586CB1	90125322CA2
7500104	44	7500104CD1	124	7500104CB1	
7500203	45	7500203CD1	125	7500203CB1	
4843802	46	4843802CD1	126	4843802CB1	4843802CA2, 90110741CA2, 90110817CA2, 90110825CA2, 90110833CA2, 90172510CA2, 90172518CA2, 90172526CA2, 90172634CA2, 90172642CA2
5877522	47	5877522CD1	127	5877522CB1	5877522CA2, 6120870CA2, 90110153CA2, 90110161CA2, 90110177CA2, 90110193CA2, 90110253CA2, 90110261CA2, 90110269CA2, 90110277CA2
617491	48	617491CD1	128	617491CB1	90109750CA2, 90109758CA2, 90109782CA2, 90109850CA2, 90109858CA2, 90109874CA2
6289901	49	6289901CD1	129	6289901CB1	90115210CA2, 90115234CA2
6817709	50	6817709CD1	130	6817709CB1	7272661CA2
6849312	51	6849312CD1	131	6849312CB1	90190001CA2, 90190009CA2, 90190025CA2, 90190033CA2, 90190041CA2, 90190109CA2, 90190117CA2, 90190125CA2
7409581	52	7409581CD1	132	7409581CB1	
7437113	53	7437113CD1	133	7437113CB1	90155830CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7500260	54	7500260CD1	134	7500260CB1	90025555CA2, 90025563CA2, 90025587CA2, 90025588CA2, 90025595CA2, 90025663CA2, 90025671CA2
7659504	55	7659504CD1	135	7659504CB1	
821165	56	821165CD1	136	821165CB1	821165CA2, 90109761CA2, 90109769CA2, 90109853CA2, 90109861CA2, 90109869CA2, 90109877CA2
7499672	57	7499672CD1	137	7499672CB1	90110796CA2, 90111017CA2, 90111033CA2
7500276	58	7500276CD1	138	7500276CB1	1218389CA2, 1875737CA2, 8168187CA2
1440723	59	1440723CD1	139	1440723CB1	90110787CA2, 90172328CA2, 90172336CA2, 90172368CA2, 90172444CA2
7479612	60	7479612CD1	140	7479612CB1	90133152CA2, 90133176CA2, 90133184CA2, 90133192CA2, 90133252CA2, 90133260CA2, 90133268CA2, 90133284CA2, 90133292CA2, 90134004CA2
1391514	61	1391514CD1	141	1391514CB1	7292618CA2
2102578	62	2102578CD1	142	2102578CB1	90132950CA2, 90132958CA2, 90132966CA2, 90132982CA2, 90132990CA2, 90133058CA2, 90133066CA2, 90133082CA2, 90133090CA2, 90133331CA2, 90197343CA2
3213122	63	3213122CD1	143	3213122CB1	3213122CA2, 6322461CA2, 90133962CA2, 90133986CA2, 90134054CA2, 90134062CA2, 90134070CA2, 90134078CA2, 90172690CA2
4326307	64	4326307CD1	144	4326307CB1	
6037749	65	6037749CD1	145	6037749CB1	6037749CA2
6285519	66	6285519CD1	146	6285519CB1	7131125CA2, 7317881CA2
70336045	67	70336045CD1	147	70336045CB1	7625761CA2, 90144712CA2
7153577	68	7153577CD1	148	7153577CB1	7153577CA2, 90132905CA2, 90132961CA2, 90132977CA2, 90132985CA2, 90133053CA2, 90133061CA2, 90133069CA2, 90133077CA2, 90133093CA2, 90133129CA2, 90133205CA2, 90133361CA2, 90133601CA2, 90133633CA2, 90133641CA2
7500299	69	7500299CD1	149	7500299CB1	
7480218	70	7480218CD1	150	7480218CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7501159	71	7501159CD1	151	7501159CB1	90189746CA2
7501932	72	7501932CD1	152	7501932CB1	
7501111	73	7501111CD1	153	7501111CB1	90023351CA2, 90023363CA2
7501113	74	7501113CD1	154	7501113CB1	
7501118	75	7501118CD1	155	7501118CB1	90023315CA2
7501128	76	7501128CD1	156	7501128CB1	90023351CA2, 90023363CA2, 90023391CA2
7501920	77	7501920CD1	157	7501920CB1	
7510325	78	7510325CD1	158	7510325CB1	90012168CA2, 90012176CA2, 90012240CA2, 90012276CA2, 90012412CA2, 90023320CA2, 90023354CA2, 90023359CA2, 90023361CA2, 90023362CA2, 90023367CA2, 90023370CA2, 90023377CA2, 90023383CA2,
					90176711CA2, 90176803CA2, 90176827CA2, 90176835CA2, 90176843CA2, 90177559CA2, 90177691CA2
7510966	79	7510966CD1	159	7510966CB1	90023363CA2, 90023395CA2, 90177567CA2
7386101	80	7386101CD1	160	7386101CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
5	2994102CD1	g1353653	3.0E-07	[fl][Strongylocentrotus purpuratus] sperm receptor for egg jelly (Moy,G.W. et al. (1996) J. Cell Biol. 133 (4), 809-817)
15	7493507CD1	g1046223	1.1E-185	[Homo sapiens] melanoma ubiquitous mutated protein (Coulie,P.G. et al. (1995) Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7976-7980)
16	3075994CD1	g3747097	9.9E-91	[Homo sapiens] C1q-related factor
25	8001939CD1	g19850565	2.0E-55	[fl][Mus musculus] NFAT activation molecule 1
28	4758058CD1	g205250	5.0E-18	[Rattus norvegicus] Ly6C antigen (Friedman,S. et al. (1990) Immunogenetics 31 (2), 104-111)
29	7499835CD1	g285971	8.5E-58	[Homo sapiens] PAP homologous protein (Itoh,T. et al. (1993) Biochim. Biophys. Acta 1172 (1-2), 184-186)
35	3040427CD1	g20799379	1.0E-66	[fl][Rattus norvegicus] neural stem cell derived neuronal survival protein precursor
43	3069586CD1	g14289183	6.7E-113	[Homo sapiens] chorein (Ueno,S. et al. (2001) Nat. Genet. 28 (2), 121-122)
44	7500104CD1	g190484	2.3E-84	[Homo sapiens] prepro salivary proline-rich protein (Maeda,N. et al. (1985) J. Biol. Chem. 260 (20), 11123-11130)
45	7500203CD1	g482909	8.5E-58	[Homo sapiens] pancreatitis-associated protein (Dusetti,N.J. et al. (1994) Genomics 19 (1), 108-114)
57	7499672CD1	g13194528	1.5E-57	[Homo sapiens] NPC-related protein NAG73
58	7500276CD1	g794071	2.6E-56	[Macaca fascicularis] epididymal secretory protein 14.6 (Perry,A.C. et al. (1995) Gene 153 (2), 291-292)
69	7500299CD1	g13543353	1.6E-68	[Homo sapiens] (BC005839) follistatin-like 3 (secreted glycoprotein)
70	7480218CD1	g13241974	9.0E-289	[Homo sapiens] CocoaCrisp
71	7501159CD1	g1747306	6.9E-156	[Mus musculus] SDR2 (Shirozu,M. et al. (1996) Genomics 37 (3), 273-280)
72	7501932CD1	g1488047	5.4E-14	[Xenopus laevis] RING finger protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
73	7501111CD1	367644|Rn.25073	7.5E-16	[Rattus norvegicus] [Receptor (signalling)] [Plasma membrane] G-protein-coupled receptor with a large extracellular domain, expressed in lung, kidney and heart Abe, J. et al. (1999) J. Biol. Chem. 274:19957-19964 Ig-hepta, a novel member of the G-protein-coupled hepta-helical receptor (GPCR) family that has immunoglobulin-like repeats in a long N- terminal extracellular domain and defines a new subfamily of GPCRs.
74	7501113CD1	367644|Rn.25073	8.5E-20	[Rattus norvegicus] [Receptor (signalling)] [Plasma membrane] G-protein-coupled receptor with a large extracellular domain, expressed in lung, kidney and heart Abe, J. *supra*
76	7501128CD1	367644|Rn.25073	2.5E-15	[Rattus norvegicus][Receptor (signalling)][Plasma membrane]G protein-coupled receptor with a large extracellular domain, expressed in lung, kidney and heart
77	7501920CD1	g5525078	5.0E-19	[Rattus norvegicus] seven transmembrane receptor (Abe,J. et al. (1999) J. Biol. Chem. 274 (28), 19957-19964)
77	7501920CD1	367644|Rn.25073	4.3E-20	[Rattus norvegicus][Receptor (signalling)][Plasma membrane] G-protein-coupled receptor with a large extracellular domain, expressed in lung, kidney and heart Abe, J. et al. *Supra*
80	7386101CD1	g458726	6.9E-27	[Homo sapiens] estrogen responsive finger protein (efp) (Inoue,S. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90 (23), 11117-11121)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1417062CD1	269	S30 S68 S222 T204 T248		Signal_cleavage: M1-G26	SPSCAN
					Signal Peptide: M25-A45, M25-A47, M25-G49	HMMER
					Non-cytosolic domains: M1-L269	TMHMMER
					Biotin repressor PF01317: Q13-R29	BLIMPS_PFAM
					Leucine zipper pattern: L155-L176	MOTIFS
					Eukaryotic putative RNA-binding region RNP-1 signature: K74-F81	MOTIFS
					TonB-dependent receptor proteins signature 1: M1-E98	MOTIFS
2	2007701CD1	127	S108 S114 T3 T70 T85	N79	Signal_cleavage: M1-W33	SPSCAN
					Signal Peptide: M11-N25, M11-A29, M11-W33, S10-W33, M11-A38, M1-Q30	HMMER
					Cytosolic domains: M1-M11, N67-W127	TMHMMER
					Transmembrane domains: L12-H34, L44-L66	TMHMMER
					Non-cytosolic domains: T35-L43	TMHMMER
3	2915695CD1	71	T64		Signal_cleavage: M1-G21	SPSCAN
					Signal Peptide: M1-Q18, M1-G21, M1-S23, M1-G25	HMMER
					Non-cytosolic domains: M1-T71	TMHMMER
4	2969449CD1	83	S61		Signal_cleavage: M1-A18	SPSCAN
					Signal Peptide: M1-A18	HMMER
					Non-cytosolic domains: M1-T83	TMHMMER
					Ribonucleotide reductase large subunit signature: A21-N72	PROFILESCAN
5	2994102CD1	306	S58 S194 T17 T162 T182 T255 T274	N45 N52 N111	Signal_cleavage: M1-G60	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Non-cytosolic domains: M1-R306	TMHMMER
6	3410251CD1	334	S70 S173 S224 S321	N147	Signal_cleavage: M1-A26	SPSCAN
					Signal Peptide: P9-A26, M1-A26, M1-A30, L7-A26	HMMER
					Non-cytosolic domains: M1-V264	TMHMMER
					Transmembrane domains: V265-L287	TMHMMER
					Cytosolic domains: T288-A334	TMHMMER
					Leucine Rich Repeat: A120-R143, A144-P167, A96-G119, L168-P191, R72-G95	HMMER_PFAM
					Leucine zipper pattern: L172-L193	MOTIFS
7	5330327CD1	950	S6 S13 S37 S48 S57 S68 S104 S114 S141 S179 S186 S340 S344 S355 S388 S396 S411 S425 S441 S566 S607 S611 S668 S680 S777 S779 S844 S855 S901 S916 T63 T110 T251 T262 T294 T309 T333 T562 T679 T793 T824 T836	N644 N853	Signal_cleavage: M1-A35	SPSCAN
					Non-cytosolic domains: M1-F950	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE PIP5K TRANSFERASE PHOSPHATIDYLINOSITOL-4-PHOSPHATE FINGER-CONTAINING PHOSPHOINOSITIDE FYVE PTDINS4P-5-KINASE 1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE	BLAST_PRODOM
8	5532048CD1	546	S17 S37 S82 S87 S186 S199 S212 S221 S237 S296 S302 S375 S402 S421 S439 S447 S488 T32 T412 T484 T493 T505 T517 T523		Signal_cleavage: M1-S58	SPSCAN
					Non-cytosolic domains: M1-I546	TMHMMER
9	56002716CD1	226	S72 S94 S108 S158 S205 T70 T118		Signal_cleavage: M1-A20	SPSCAN
					Non-cytosolic domains: M1-I226	TMHMMER
					Signal Peptide: M1-T17, M1-A18, M1-A21, M1-V23, M1-A20	HMMER
10	60129797CD1	130	S40 S108 T46		Signal_cleavage: M1-A23	SPSCAN
					Signal Peptide: M1-V21, M1-A23, M1-S24, M1-G32, M1-T26	HMMER
					Non-cytosolic domains: M1-S130	TMHMMER
11	6246243CD1	195	S4 S33 S40 S41 S114 S141 S146 S161 T65 T155	N62 N159	Signal_cleavage: M1-L24	SPSCAN
					Signal Peptide: M1-R25, M1-L24, M1-P26	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Non-cytosolic domains: M1-Q195	TMHMMER
12	6804755CD1	112	S7 S42		Signal_cleavage: M1-A62	SPSCAN
					Signal Peptide: M8-G27, M8-Q29, M8-A33, M1-A33, M8-P30	HMMER
					Cytosolic domains: M1-G112	TMHMMER
					Aldo/keto reductase family signatures: M8-P73	PROFILESCAN
13	6856852CD1	107	S22 S35		Signal_cleavage: M1-G16	SPSCAN
					Signal Peptide: M1-G16, M1-A18, M1-S20, M1-S22, M1-R24, M1-S20	HMMER
					Non-cytosolic domains: M1-L107	TMHMMER
14	7482027CD1	221	S100 T112 T155 T211	N91	Signal_cleavage: M1-G14	SPSCAN
					Signal Peptide: M1-A15, M1-W19, M1-G20	HMMER
					Cytosolic domains: C210-Q221	TMHMMER
					Non-cytosolic domains: M1-Q186	TMHMMER
					Transmembrane domains: A187-A209	TMHMMER
15	7493507CD1	642	S5 S22 S50 S61 S78 S147 S165 S167 S240 S297 S306 S315 S318 S363 S372 S390 S393 S439 S472 S527 S531 S601 S622 T44 T171 T270 T274 T545 T549 T579 T610 Y615	N66 N142 N162	Signal_cleavage: M1-S37	SPSCAN
					Non-cytosolic domains: M1-R642	TMHMMER
					Cell attachment sequence: R72-D74	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	3075994CD1	238	S15 S143 T224		Signal_cleavage: M1-S15	SPSCAN
					Signal Peptide: M1-S15, M1-G18, M1-A20	HMMER
					Non-cytosolic domains: M1-D238	TMHMMER
					C1q domain: A111-I235	HMMER_PFAM
					C1q domain proteins BL01113: G73-G99, V125-V160, D194-K213, S228-P237	BLIMPS_BLOCKS
					Complement C1Q domain signature PR00007: P119-K145, F146-G165, D194-D215, K226-Y236	BLIMPS_PRINTS
					PRECURSOR SIGNAL COLLAGEN REPEAT HYDROXYLATION GLYCOPROTEIN CHAIN PLASMA EXTRACELLULAR MATRIX PD002992: R118-I235	BLAST_PRODOM
					COLLAGEN ALPHA PRECURSOR CHAIN REPEAT SIGNAL CONNECTIVE TISSUE EXTRACELLULAR MATRIX PD000007: G36-G99	BLAST_PRODOM
					SIMILAR TO CUTICULAR COLLAGEN PD067228: G18-P102	BLAST_PRODOM
					PRECURSOR SIGNAL COLLAGEN ALPHA 3IX CHAIN EXTRACELLULAR MATRIX CONNECTIVE TISSUE PD028299: G36-G98	BLAST_PRODOM
					C1Q DOMAIN DM00777	BLAST_DOMO
					P02746|70-250: G58-D238	BLAST_DOMO
					P23206|477-673: R63-P237	BLAST_DOMO
					S23297|465-674: P51-I234	BLAST_DOMO
					S49158|70-253: G58-D238	BLAST_DOMO
					C1q domain signature: F128-Y158	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	2378119CD1	113	S49 S72 S90 T10 T25 Y35		signal_cleavage: M1-S49	SPSCAN
18	2987418CD1	97	S76 T68		Signal Peptide: M1-G23	HMMER
					signal_cleavage: M25-N92	SPSCAN
					inside:M1-T33	TMHMMER
					TMhelix:F34-I56	TMHMMER
					outside:L57-K97	TMHMMER
19	4223862CD1	147	S68 S80		signal_cleavage: M1-G36	SPSCAN
					inside:M1-H93	TMHMMER
					TMhelix:L94-L116	TMHMMER
					outside:G117-N147	TMHMMER
20	6046406CD1	95	S45 S91		signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A22 M1-A24	HMMER
21	6743529CD1	76		N41	signal_cleavage: M1-G20	SPSCAN
					Signal Peptide: M1-C19	HMMER
22	7283809CD1	154			signal_cleavage: M1-S26	SPSCAN
					Signal Peptide: M2-S19, M2-S24, M2-S26, M2-G31	HMMER
23	7637563CD1	160	S60 S79 T49 T61 T136 T140		signal_cleavage: M1-G24	SPSCAN
					Signal Peptide: M1-G24	HMMER
24	7663814CD1	72	S19 S50	N41 N55	signal_cleavage: M1-P21	SPSCAN
					Signal Peptide: M1-P21, M1-P23, M1-S24	HMMER
25	8001939CD1	270	S64 S92 S210 T36 T97 T139 T199 T228 T243	N107	signal_cleavage: M1-G42	SPSCAN
					Signal Peptide: M1-G42	HMMER
					outside:M1-K163	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					TMhelix:L164-W186	TMHMMER
					inside:N187-L270	TMHMMER
26	8191019CD1	121			signal_cleavage: M1-P18	SPSCAN
					Signal Peptide: M1-P18, M1-P21	HMMER
					F-actin capping protein beta subunit signature: G17-P92	PROFILESCAN
27	919788CD1	181	S87		signal_cleavage: M1-A41	SPSCAN
28	4758058CD1	120	S55		signal_cleavage: M1-P22	SPSCAN
					Signal Peptide: M7-S21, M7-P24, M7-G26, M7-C29, M1-G26, M1-C29	HMMER
					u-PAR/Ly-6 domain: M1-V60, S83-L120	HMMER_PFAM
					Ly-6 / u-PAR domain proteins BL00983: L12-L20, Q23-C32, A76-N91	BLIMPS_BLOCKS
					LY-6 / U-PAR DOMAIN	BLAST_DOMO
					DM02129|P35460|1-133: M1-L120	BLAST_DOMO
					DM02129|I48639|1-134: M1-L119	BLAST_DOMO
					DM02129|P09568|1-130: M1-L120	BLAST_DOMO
					DM02129|P05533|1-133: M1-L120	BLAST_DOMO
29	7499835CD1	129	S35 S57 S60 S67 S77 T29		signal_cleavage: M1-G26	SPSCAN
					Signal Peptide: M13-T29, S9-G26, M5-G26, M5-T29, M1-E28, M1-G26, M5-V24	HMMER
					Lectin C-type domain: T29-F127	HMMER_PFAM
					C-type lectin domain proteins BL00615: C51-E68, W112-C125	BLIMPS_BLOCKS
					C-type lectin domain signature and profile: D80-K128	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PRECURSOR SIGNAL PROTEIN LECTIN REG LITHOSTATHINE REGENERATING INFLAMMATORY RESPONSE ACUTE PD149843: L21-D65	BLAST_PRODOM
					C-TYPE LECTIN	BLAST_DOMO
					DM00035|Q06141|33-172: G66-F127, L33-D65	BLAST_DOMO
					DM00035|P35230|33-172: G66-F127, L33-D65	BLAST_DOMO
					DM00035|P23132|33-172: D65-F127, L33-D65	BLAST_DOMO
					DM00035|S54979|33-171: G66-F127, L33-D65	BLAST_DOMO
					C-type lectin domain signature: C100-C125	MOTIFS
30	2484647CD1	101	S82		signal_cleavage: M1-A36	SPSCAN
					Signal Peptide: M1-A19, M1-P21	HMMER
31	2587034CD1	83	T76		signal_cleavage: M1-A28	SPSCAN
					Signal Peptide: M1-G21	HMMER
					Cytosolic domain: I33-F83	TMHMMER
					Transmembrane domain: F10-L32	TMHMMER
					Non-cytosolic domain: M1-Y9	TMHMMER
					Pancreatic ribonuclease family signature: N5-I61	PROFILESCAN
					Indole-3-glycerol phosphate synthase signature: L25-G77	PROFILESCAN
					Leucine zipper pattern: L11-L32	MOTIFS
32	2702991CD1	172	T14		signal_cleavage: M1-A35	SPSCAN
33	2744736CD1	168	S100 S147		signal_cleavage: M1-A56	SPSCAN
					Signal Peptide: M24-P43	HMMER
34	2915475CD1	83	T37		signal_cleavage: M1-A18	SPSCAN
					Signal Peptide: M1-A18, M1-E20, M1-T22, M1-G26	HMMER
					Cytosolic domain: T72-T83	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Transmembrane domain: L49-F71	TMHMMER
					Non-cytosolic domain: M1-H48	TMHMMER
35	3040427CD1	167	S89 S134 T23 T66		signal_cleavage: M1-A47	SPSCAN
					Signal Peptide: M24-A39, M24-C41, M24-G44, M24-A47, M22-B49, L28-A47, M22-A47	HMMER
					EF-hand clacium-binding domain protein BL00018: D150-F162	BLIMPS_BLOCKS
					Laminin-type EGF-like (LE) domain protein BL01248: C1061-C1354	BLIMPS_BLOCKS
					CALMODULIN REPEAT DM00011|S0027|29-74: T116-A163	BLAST_DOMO
					EF-HAND CALCIUM-BINDING DOMAIN DM00256|S0027|1-27: M88-S115	BLAST_DOMO
					Binding-protein-dependent transport systems inner membrane component. signature: M1-R29	MOTIFS
					EF-hand calcium-binding domain: D102-L114, D150-F162	MOTIFS
36	7499722CD1	195	S100 S161 S165 S182 T76 T122	N123 N192	signal_cleavage: M1-A28	SPSCAN
					Signal Peptide: M1-G21	HMMER
					Cytosolic domain: I33-L195	TMHMMER
					Transmembrane domain: F10-L32	TMHMMER
					Non-cytosolic domain: M1-Y9	TMHMMER
					Leucine zipper pattern: L11-L32	MOTIFS
37	6776909CD1	89	S71		signal_cleavage: M1-G27	SPSCAN
					Signal Peptide: M1-G27, M1-A29	HMMER
38	7280438CD1	136	S51 S115		signal_cleavage: M1-A16	SPSCAN
					Signal Peptide: M1-A16, M1-G18, M1-G20	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
39	7499809CD1	420	S90 S180 S274 S383 S413 T46 T319 T336 T353 T388	N210	signal_cleavage: M1-C18	SPSCAN
					Signal Peptide: M1-C18, M1-V20, M1-G23, M1-G27	HMMER
					Regulator of chromosome condensation (RCC1) signature 2: V140-L150	MOTIFS
					Leucine zipper pattern: L153-L174, L160-L181, L167-L188, L303-L324	MOTIFS
40	7499921CD1	667	S68 S144 S186 S229 S230 S231 S245 S339 S471 S475 S494 S627 S644 T89 T102 T221 T411 T603	N659	signal_cleavage: M1-G46	SPSCAN
					Flavodoxin: G78-V126	HMMER_PFAM
					Cytosolic domain: M1-I19	TMHMMER
					Transmembrane domain: N20-I42	TMHMMER
					Non-cytosolic domain: K43-C667	TMHMMER
41	2705858CD1	83			signal_cleavage: M1-S18	SPSCAN
					Signal Peptide: M1-S18, M1-T20, M1-A23, M1-R25	HMMER
42	3069892CD1	80	S7 S46 S47 S62 T72		signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A24, M20-A51	HMMER
43	3069586CD1	367	S54 S62 S82 S87 S326 S332 T27 T36 T110 T112		signal_cleavage: M1-P29	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN VACUOLAR SORTINGASSOCIATED VPS13 TIPC T08G11.1 PD025730: V2-E360	BLAST_PRODOM
44	7500104CD1	154	S15 S47		signal_cleavage: M1-A16	SPSCAN
					Signal Peptide: M1-A16, M1-D18	HMMER
					SALIVARY ACIDIC PROLINE RICH PHOSPHOPROTEIN 1/2 PRECURSOR PRP1/PRP 3 PRP2/PRP4 PIFF/PIFS PROTEIN A/PROTEIN C CONTAINS: PEPTIDE PC REPEAT SALIVA SIGNAL PAROTID GLAND PHOSPHORYLATION PD054888: M1-D55	BLAST_PRODOM
					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI ANCHOR BRAIN MAJOR PD001091: R34-Q154	BLAST_PRODOM
					COLLAGEN ALPHA PRECURSOR CHAIN REPEAT SIGNAL CONNECTIVE TISSUE EXTRACELLULAR MATRIX PD000007: G66-P153	BLAST_PRODOM
					TRACHEAL COLONIZATION FACTOR PRECURSOR SIGNAL SARCAL UMENIN CALCIUM BINDING GLYCOPROTEIN ALTERNATIVE SPLICING PD136752: D18-G150	BLAST_PRODOM
					PROLINE-RICH PROTEIN	BLAST_DOMO
					DM01369|P02810|86-164: P74-P153	BLAST_DOMO
					DM01281|P04280|17-124: Q42-G141	BLAST_DOMO
					DM03894|A39066|1-159: M1-Q154	BLAST_DOMO
					DM01281|P04280|212-315: G56-G146	BLAST_DOMO
45	7500203CD1	129	S35 S57 S60 S67 S77 T29		signal_cleavage: M1-G26	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M13-T29, M5-G26, M5-T29, M1-E28, M1-G26, M5-V24	HMMER
					Lectin C-type domain: T29-F127	HMMER_PFAM
					C-type lectin domain proteins BL00615: C51-E68, W112-C125	BLIMPS_BLOCKS
					C-type lectin domain signature and profile: D80-K128	PROFILESCAN
					PRECURSOR SIGNAL PROTEIN LECTIN REG LITHOSTATHINE REGENERATING INFLAMMATORY RESPONSE ACUTE PD149843: L21-D65	BLAST_PRODOM
					C-TYPE LECTIN	BLAST_DOMO
					DM00035|Q06141|33-172: G66-F127, L33-D65	BLAST_DOMO
					DM00035|P35230|33-172: G66-F127, L33-D65	BLAST_DOMO
					DM00035|P23132|33-172: D65-F127, L33-D65	BLAST_DOMO
					DM00035|S54979|33-171: G66-F127, L33-D65	BLAST_DOMO
					C-type lectin domain signature: C100-C125	MOTIFS
46	4843802CD1	116	S65 S66 S72 T89		signal_cleavage: M1-G23	SPSCAN
					Signal Peptide: M1-G23, M1-S30, M1-A25, M1-G28	HMMER
47	5877522CD1	84	T14		signal_cleavage: M1-R16	SPSCAN
					Signal Peptide: M1-R16	HMMER
48	617491CD1	83	S26 S40 T39	N37	signal_cleavage: M1-S28	SPSCAN
					Signal Peptide: M1-A23, M1-S30, M1-S28	HMMER
49	6289901CD1	133	S40 S72 S89 S97		signal_cleavage: M1-G38	SPSCAN
					Myelin proteolipid protein signatures: D10-A61	PROFILESCAN
50	6817709CD1	117	S52 T28 T30		signal_cleavage: M1-A15	SPSCAN
					Signal Peptide: M1-A15	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
51	6849312CD1	99	S23 S80 T69	N19	signal_cleavage: M1-S23	SPSCAN
					Signal Peptide: M1-S23	HMMER
52	7409581CD1	114	S38 S58 S61 T35		signal_cleavage: M1-G34	SPSCAN
					Signal Peptide: M1-G34	HMMER
					Cytosolic domain: Q33-G114	TMHMMER
					Transmembrane domain: P15-S32	TMHMMER
					Non-cytosolic domain: M1-Q14	TMHMMER
53	7437113CD1	699	S99 S101 S152 S236 S280 S284 S312 S327 S335 S441 S459 S492 S501 S502 S590 S631 S636 T54 T67 T294 T298 T375 T398 T399 T403 T475 T545 T547 T604 T675 T684	N97 N333 N352 N490 N524 N613	signal_cleavage: M1-S33	SPSCAN
54	7500260CD1	144	S115	N53	signal_cleavage: M1-A51	SPSCAN
55	7659504CD1	382	S17 S109 S195 S279 T105 T139 T175 T227	N154 N263	signal_cleavage: M1-G35	SPSCAN
					Cell attachment sequence: R32-D34	MOTIFS
					Leucine zipper pattern: L329-L350	MOTIFS
56	821165CD1	93			signal_cleavage: M1-Q46	SPSCAN
					Signal Peptide: M1-A17, M1-P19, M1-S24, M1-L25, M1-P18, M1-S23	HMMER
57	7499672CD1	110	S65		signal_cleavage: M1-C61	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
58	7500276CD1	115	S33 S85	N99	signal_cleavage: M1-A19	SPSCAN
					Signal Peptide: M1-A16, M1-A19, M1-P21, M1-Q23, M1-F24	HMMER
					E1 family: A6-I111	HMMER_PFAM
					PRECURSOR SIGNAL ALLERGEN PROTEIN MITE SECRETORY E1 POLYMORPHISM DER II PD008264: S31-I111, M1-C27	BLAST_PRODOM
59	1440723CD1	161	S133		signal_cleavage: M1-G47	SPSCAN
					Signal Peptide: M1-V21	HMMER
60	7479612CD1	88	T38		signal_cleavage: M1-T38	SPSCAN
					Signal Peptide: M1-V15	HMMER
61	1391514CD1	79	S18 S75		signal_cleavage: M1-C23	SPSCAN
					Signal Peptide: M1-C23	HMMER
62	2102578CD1	76	S18 S25 T63		signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A24	HMMER
63	3213122CD1	116	S29 S67		signal_cleavage: M1-Q15	SPSCAN
					Cytochrome c family heme-binding site signature: C106-S111	MOTIFS
64	4326307CD1	558	S21 S82 S128 S158 S237 S242 S244 S296 S300 S318 S329 S339 S365 S386 S484 S543 T135 T226 T346 T448 T449 T476 Y49 Y143	N86 N262 N327	signal_cleavage: M1-S21	SPSCAN
					Signal Peptide: M1-S21	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
65	6037749CD1	155	T36 T77 T109 T124 T143 Y151	N67 N98 N122	signal_cleavage: M1-A25	SPSCAN
					Signal Peptide: M7-S23, M1-A25, M7-A25	HMMER
66	6285519CD1	77	T7		signal_cleavage: M1-C32	SPSCAN
					Signal Peptide: L14-C32, M8-G31, M8-C32	HMMER
67	70336045CD1	240	S11 S56 S76 S90 S149 S178 S192 S204 S205 T238	N202	signal_cleavage: M1-A44	SPSCAN
					EF hand: D96-L124, G132-A160	HMMER_PFAM
68	7153577CD1	101	S32 S36 S47 S52		signal_cleavage: M1-A33	SPSCAN
					Signal Peptide: M1-A20, M1-S26	HMMER
69	7500299CD1	129	S121 T112	N81	signal_cleavage: M1-S26	SPSCAN
					Signal Peptide: M1-S26, M1-A20	HMMER
					Kazal-type serine protease inhibitor domain: C66-C109	HMMER_PFAM
					Osteonectin domain signatures: Q46-C88	PROFILESCAN
					KAZAL PROTEINASE INHIBITOR DM00123|P50291|186-238: S55-C109	BLAST_DOMO
70	7480218CD1	500	S64 S81 S98 S136 S176 S207 S245 S278 S279 S367 T4 T30 T58 T273 T398 T471 Y249	N28	signal_cleavage: M1-A20	SPSCAN
					SCP-like extracellular protein: Q63-G214	HMMER_PFAM
					Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins BL01009: M86-C103, H133-Y146, T166-C186, V200-H215	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Allergen V5/Tpx-1 family signature PR00837: C165-C181, Y201-G214, M86-L104, H133-Y146	BLIMPS_PRINTS
					Venom allergen 5 signature PR00838: M86-L104, T131-Y146, V164-I183	BLIMPS_PRINTS
					PROTEIN PRECURSOR SIGNAL PATHOGENESISRELATED ANTIGEN ALLERGEN VENOM MULTIGENE FAMILY AG5 PD000542: P78-G214	BLAST_PRODOM
					FSG 120K CYSRICH PROTEIN GLYCOPROTEIN EGFLIKE DOMAIN PD128352: G53-G232	BLAST_PRODOM
					EXTRACELLULAR PROTEINS SCP/TPX-1/AG5/PR-1/SC7 DM00332|P48060|1-175: I57-Y218	BLAST_DOMO
					EXTRACELLULAR PROTEINS SCP/TPX-1/AG5/PR-1/SC7 DM00332|P54108|11-182: T58-W212	BLAST_DOMO
					EXTRACELLULAR PROTEINS SCP/TPX-1/AG5/PR-1/SC7 DM00332|P16562|9-180: Q63-N211	BLAST_DOMO
					EXTRACELLULAR PROTEINS SCP/TPX-1/AG5/PR-1/SC7 DM00332|P35778|12-207: T58-P217	BLAST_DOMO
					Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2: Y201-W212	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
71	7501159CD1	402	S116 S121 S194 S305 S358 T57 T95 T106 T145 T172 T225 T276 T289 Y129	N138 N309 N322	signal_cleavage: M1-N22	SPSCAN
					Signal Peptide: M1-V20	HMMER
					Reeler domain: S31-K156	HMMER_PFAM
					SDR2 PROTEIN PD139571: Q164-S367	BLAST_PRODOM
					PROTEIN SDR2 BASIC HEMOLYMPH PRECURSOR SIGNAL PD035283: P24-S163	BLAST_PRODOM
72	7501932CD1	363	S27 S193 S293	N178	signal_cleavage: M1-S31	SPSCAN
					SPRY domain: A227-P349	HMMER_PFAM
					PF00622 Domain in Spla and the Ryanodine Receptors (SPRY domain)	BLIMPS_PFAM
					PROTEIN FINGER MIDLINE ZINC FINGER RING STONUSTOXIN PUTATIVE TRANSCRIPTION FACTOR XPRF PD002421: K171-L343	BLAST_PRODOM
					RFP TRANSFORMING PROTEIN DM01944|I49642|513-634: G228-C346	BLAST_DOMO
					RFP TRANSFORMING PROTEIN DM01944|A49656|508-630: G228-C346	BLAST_DOMO
73	7501111CD1	221	S115 T13 T30 T55 T104 T155	N139 N168	Signal_cleavage: M1-G17	SPSCAN
					Signal Peptide: M1-G17, M1-G19, M1-G20, M1-G23	HMMER
74	7501113CD1	267	S115 S219 S225 S248 T13 T30 T55 T104 T155	N139 N168 N205	Signal_cleavage: M1-G17	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M1-G17, M1-G19, M1-G20, M1-G23	HMMER
75	7501118CD1	236	S115 S225 T13 T30 T55 T104 T155	N139 N168	signal_cleavage: M1-G17	SPSCAN
					Signal Peptide: M1-G17, M1-G19, M1-G20, M1-G23	HMMER
76	7501128CD1	221	S115 T13 T30 T55 T104 T155	N139 N168	signal_cleavage: M1-G17	SPSCAN
					Signal Peptide: M1-G17, M1-G19, M1-G20, M1-G23	HMMER
					PHD-finger. PF00628: C105-P119	BLIMPS_PFAM
77	7501920CD1	410	S96 S200 S206 S229 S247 S283 S350 S357 S401 T13 T36 T85 T136 T265 T351	N120 N149 N186 N263 N291 N298 N310 N335 N349	signal_cleavage: M1-G17	SPSCAN
					Signal Peptide: M1-G17, M1-G19, M1-G20, M1-G23	HMMER
78	7510325CD1	67			signal_cleavage: M1-P50	SPSCAN
					Sigma-54 interaction domain signatures and profile: L9-R56	PROFILESCAN
					Eukaryotic mitochondrial porin signature: M3-F66	PROFILESCAN
79	7510966CD1	49				
80	7386101CD1	495	S166 S181 S325 S425 T118 T153	N310	SPRY domain: A359-P481	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C12-C50	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Zinc finger, C3HC4 type (RING finger), signature: L6-R61	PROFILESCAN
					Zinc finger C3HC4 type BL00518: C27-C35	BLIMPS_BLOCKS
					PROTEIN FINGER MIDLINE ZINCFINGER RING STONUSTOXIN PUTATIVE TRANSCRIPTION FACTOR XPRF PD002421: V316-L475	BLAST_PRODOM
					RFP TRANSFORMING PROTEIN DM01944 I49642|513-634: G360-C478 A49656|508-630: G360-C478	BLAST_DOMO
					ZINC FINGER, C3HC4 TYPE, DM00063 A49656|6-55: L6-R51 A43906|137-187: E8-E52	BLAST_DOMO
					Zinc finger, C3HC4 type (RING finger), signature: C27-I36	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
81/1417062CB1/ 1146	1-233, 1-528, 1-1146, 10-618, 17-281, 17-567, 20-285, 20-664, 23-316, 23-524, 25-874, 28-335, 127-578, 184-853, 742-1145, 786-984, 786-1016, 796-1063
82/2007701CB1/ 1043	1-81, 18-489, 18-702, 187-702, 315-702, 363-656, 364-702, 491-663, 581-702, 589-1043
83/2915695CB1/ 1684	1-704, 31-667, 31-733, 31-955, 32-1684, 44-578, 59-689, 60-629, 73-785, 74-785, 97-415, 191-447, 209-1192, 210-936, 227-497, 272-936, 274-1180, 280-785, 392-646, 492-524, 492-539, 492-542, 492-548, 492-554, 492-559,
	492-569, 492-572, 492-584, 492-599, 492-614, 499-785, 522-600, 522-644, 529-644, 552-638, 552-644, 559-644, 582-644, 589-644, 612-644, 619-644, 709-964, 711-964, 715-964, 725-962, 821-1120, 836-1122, 1269-1373, 1269-1377, 1271-1366, 1313-1377, 1369-1422
84/2969449CB1/ 1584	1-755, 39-642, 47-857, 76-871, 217-922, 224-748, 228-736, 240-978, 252-633, 252-922, 252-977, 280-758, 283-995, 317-839, 319-839, 321-772, 321-967, 322-767, 327-799, 330-790, 330-814, 335-820, 342-820, 351-894,
	402-975, 406-939, 408-944, 431-980, 462-1062, 471-1060, 547-980, 556-1410, 560-1217, 560-1272, 566-1235, 572-1079, 580-1115, 598-1297, 625-1584, 657-1584, 666-1311, 670-1580, 687-1340, 717-1583, 790-1584,
	810-1270, 857-1584, 869-1584, 918-1329, 924-1584, 1074-1362, 1074-1584, 1097-1584, 1109-1584, 1121-1356, 1135-1584, 1141-1584, 1142-1584, 1185-1584, 1274-1583
85/2994102CB1/ 1490	1-274, 1-1490, 19-684, 27-326, 27-434, 38-607, 39-709, 49-357, 641-884, 641-1199, 846-1114, 863-1474, 866-1089, 866-1090, 914-989, 990-1490, 1372-1476
86/3410251CB1/ 1418	1-233, 1-754, 310-494, 310-797, 382-635, 429-814, 646-1073, 655-1108, 667-1074, 673-1096, 692-1073, 700-1102, 703-1073, 717-1073, 766-1093, 770-1048, 771-1073, 790-1189, 790-1418, 864-1033
87/5330327CB1/ 3485	1-296, 1-467, 75-710, 75-717, 75-792, 334-868, 399-922, 456-790, 805-1511, 819-2112, 821-1335, 839-1335, 845-1081, 857-1345, 858-1299, 868-1345, 933-1345, 1095-1335, 1138-1409, 1339-1878, 1343-1857, 1343-1973,
	1346-1734, 1346-1751, 1346-1786, 1346-1794, 1346-1867, 1346-1868, 1346-1881, 1346-1882, 1346-1898, 1346-1912, 1346-1933, 1348-1944, 1424-1976, 1444-1948, 1550-2410, 1571-1938, 1571-2161, 1573-2173, 1701-2410,
	1763-2141, 1772-2153, 1779-2410, 1806-2410, 1842-2410, 1854-2488, 1861-2410, 1875-2410, 1877-2112, 1880-2410, 1973-2591, 1991-2410, 2111-2488, 2165-2410, 2168-2410, 2504-2739, 2504-2763, 2504-2974, 2571-3415, 2641-3372, 2665-2927, 2781-2931, 2859-3485

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
88/5532048CB1/ 3427	1-772, 1-834, 1-3044, 13-220, 108-780, 382-599, 416-693, 416-878, 416-1064, 434-870, 434-953, 491-758, 594-1129, 602-1084, 619-985, 619-1529, 638-942, 642-1096, 705-1243, 718-912, 779-1028, 783-1087, 788-1376, 836-1657,
	955-1176, 976-1560, 976-1601, 981-1295, 983-1215, 983-1380, 983-1554, 986-1248, 986-1437, 1069-1210, 1101-1622, 1101-1797, 1101-1941, 1102-1663, 1103-1659, 1189-1673, 1200-1923, 1242-1856, 1251-1990, 1260-1726,
	1271-1473, 1284-1924, 1292-1823, 1299-1503, 1324-1580, 1335-2003, 1336-1959, 1353-1891, 1369-2003, 1387-1512, 1395-1988, 1457-1950, 1480-1763, 1499-1584, 1541-2081, 1541-2083, 1549-1758, 1609-1888, 1623-2003,
	1652-1958, 1652-1994, 1654-2332, 1656-2004, 1658-2327, 1739-1999, 1773-1943, 1825-2081, 1926-2093, 1946-1987, 2036-2451, 2061-2451, 2238-2661, 2376-2451, 2395-2448, 2530-3044, 2691-3033, 2753-3427, 2820-3427, 2863-3427, 2909-3427, 2964-3216, 2964-3402
89/56002716CB1/ 1438	1-27, 1-398, 1-457, 1-495, 1-496, 1-575, 1-664, 1-767, 1-780, 1-811, 5-766, 15-670, 34-885, 45-860, 56-669, 210-911, 576-1357, 663-1434, 786-1333, 840-1356, 849-1356, 886-1438, 1413-1434
90/60129797CB1/ 1710	1-453, 2-1710, 27-454, 111-566, 134-382, 163-411, 279-1035, 444-1008, 471-1152, 543-1121, 572-1334, 603-1334, 657-1334, 698-907, 708-1274, 725-983, 725-996, 725-1232, 749-997, 751-959, 756-1030, 756-1320,
	764-1050, 769-1055, 793-1047, 811-972, 812-1116, 819-1066, 829-1112, 829-1114, 829-1306, 829-1353, 829-1366, 829-1368, 830-1407, 831-1473, 833-1395, 834-1507, 836-1356, 837-1352, 839-1101, 843-1667,
	851-1169, 859-1128, 861-1092, 862-1122, 871-1460, 872-1079, 876-1449
91/6246243CB1/753	1-100, 1-640, 3-577, 7-381, 154-628, 154-655, 181-753, 236-753, 272-736, 275-732, 341-731, 342-652, 347-734, 368-732, 404-734, 411-734, 645-736

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
92/6804755CB1/ 1780	1-220, 1-258, 1-369, 1-420, 1-499, 1-513, 1-526, 1-649, 3-634, 7-627, 32-844, 87-420, 202-692, 256-723, 295-1137, 312-892, 379-1033, 409-918, 436-1103, 481-1212, 510-1091, 551-1100, 630-1419, 658-1019, 680-1183, 693-1097, 699-782, 699-794, 699-798, 699-802, 699-945, 700-761, 700-1200, 703-759, 703-761, 703-790, 703-794, 703-802, 703-803, 703-839, 703-848, 703-989, 703-1014, 703-1033, 703-1039, 704-792, 705-839, 705-1014, 706-802, 707-780, 707-802, 707-902, 713-945, 716-765, 716-789, 716-902, 716-931, 716-945, 716-947, 716-989, 717-800, 717-945, 717-975, 717-977, 717-987, 717-989, 718-787, 718-945, 719-802, 719-935, 722-802, 722-975, 723-934, 730-792, 730-802, 731-979, 734-1021, 738-802, 747-878, 747-887, 747-1058, 747-1065, 747-1077, 750-1077, 757-989, 760-989, 760-1033, 761-1033, 763-902, 763-1192, 770-1192, 778-1215, 784-1033, 786-1077, 810-900, 810-1005, 810-1077, 811-1005, 811-1077, 816-945, 816-1023, 820-989, 820-1059, 833-1077, 834-1172, 839-1077, 850-979, 858-943, 858-979, 858-989, 858-992, 858-1065, 858-1077, 860-902, 860-1077, 861-989, 869-945, 869-1077, 872-950, 876-1054, 881-1033, 892-945, 892-987, 892-1077, 904-987, 904-1067, 906-985, 912-1062, 912-1077, 916-987,
	919-1077, 924-1719, 927-1023, 927-1031, 931-1031, 935-1009, 935-1010, 935-1031, 935-1075, 938-1010, 945-1077, 946-1052, 946-1065, 946-1077, 955-1026, 958-1037, 974-1077, 975-1049, 975-1057, 975-1063,
	975-1065, 975-1075, 976-1064, 999-1077, 1002-1160, 1002-1405, 1003-1075, 1010-1562, 1032-1077, 1034-1077, 1191-1701, 1331-1401, 1336-1780
93/6856852CB1/580	1-573, 1-580, 15-577
94/7482027CB1/731	1-394, 66-495, 70-726, 292-731, 300-730, 337-729, 419-729

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
95/7493507CB1/ 2758	1-806, 10-504, 10-2758, 24-648, 29-200, 29-606, 32-324, 127-421, 204-705, 579-1233, 657-1327, 662-1313, 684-1219, 726-1402, 732-1351, 739-1369, 741-1402, 767-1419, 780-1099, 833-1494, 836-1384, 850-1351, 851-1406, 855-1457, 874-1521, 875-1418, 910-1506, 934-1486, 963-1606, 995-1430, 1013-1621, 1059-1456, 1084-1745, 1102-1309, 1110-1382, 1111-1785, 1158-1761, 1202-1771, 1208-1323, 1212-1735, 1227-1787, 1229-1451, 1229-1474, 1251-1874, 1251-1909, 1283-1876, 1294-1552, 1325-1795, 1326-2043, 1331-1830, 1347-1726, 1350-2000, 1362-1627, 1373-2077, 1392-1705, 1425-1475, 1427-1686, 1427-1982, 1433-2054, 1444-1693, 1450-1998, 1470-1733, 1470-1739, 1470-1935, 1471-1774, 1510-2288, 1512-2129, 1520-1795, 1530-2141, 1541-1832, 1546-2257, 1550-2191, 1554-2315, 1568-1847, 1579-1835, 1580-2130, 1582-2110, 1616-2308, 1617-1862, 1617-2165, 1627-2238, 1633-2152, 1647-1918, 1648-1913, 1677-2301, 1703-2250, 1703-2339, 1709-2293, 1730-2248, 1734-2347, 1737-1985, 1762-2086, 1766-2298, 1773-2034, 1777-2018, 1783-2602, 1785-2352, 1808-2320, 1815-2488, 1843-2443, 1870-2096, 1870-2347, 1893-2250, 1895-2203, 1943-2135, 1943-2191, 1943-2249, 1943-2501, 1962-2212,
	1962-2239, 1972-2239, 1977-2561, 1982-2649, 2005-2529, 2012-2645, 2020-2486, 2039-2278, 2039-2293, 2039-2602, 2063-2700, 2066-2696, 2068-2664, 2082-2707, 2087-2339, 2121-2491, 2127-2745, 2131-2758, 2160-2712,
	2170-2717, 2170-2723, 2171-2712, 2174-2722, 2181-2712, 2182-2712, 2211-2721, 2222-2712, 2234-2551, 2240-2712, 2255-2721, 2267-2640, 2271-2721, 2275-2716, 2278-2679, 2283-2714, 2285-2538, 2288-2706, 2288-2721, 2296-2722,
	2305-2593, 2320-2721, 2328-2722, 2340-2721, 2351-2721, 2360-2718, 2373-2636, 2385-2499, 2388-2544, 2425-2715, 2439-2719, 2468-2715, 2471-2716, 2471-2718, 2499-2758, 2526-2718, 2535-2670, 2535-2718, 2587-2709, 2587-2721, 2651-2719
96/3075994CB1/ 1383	1-1361, 490-531, 550-816, 555-1189, 659-1298, 757-1043, 778-1295, 815-1361, 933-1383
97/2378119CB1/826	1-230, 28-560, 46-343, 49-288, 50-298, 50-561, 51-342, 52-557, 53-224, 56-317, 56-338, 56-545, 59-552, 59-558, 59-564, 59-566, 69-382, 69-559, 70-191, 70-392, 71-361, 72-513, 72-522, 72-526, 74-297, 74-338, 76-522, 76-565, 77-354,
	77-355, 78-222, 78-351, 78-377, 80-342, 80-343, 80-356, 80-364, 82-347, 83-340, 83-539, 83-576, 84-409, 85-550, 87-569, 89-336, 89-562, 91-362, 92-549, 93-378, 93-445, 93-466, 96-382, 97-383, 102-546, 106-345, 106-383, 106-560,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	109-368, 110-358, 110-362, 110-549, 111-305, 111-329, 111-358, 111-377, 111-378, 111-402, 111-403, 111-560, 111-561, 111-565, 112-355, 113-349, 113-610, 114-553, 115-375, 116-526, 117-561, 120-380, 128-546, 128-549, 129-547,
	132-550, 132-568, 139-401, 144-552, 151-550, 154-550, 169-533, 182-551, 203-506, 214-546, 225-552, 253-374, 253-561, 254-491, 275-525, 290-826, 309-491, 332-560, 332-608, 344-560, 360-554
98/2987418CB1/ 1025	1-279, 1-700, 1-1025, 107-959, 422-570, 422-856, 473-1006, 620-1025, 732-1025, 782-1025
99/4223862CB1/ 1223	1-893, 380-593, 380-907, 398-908, 398-1221, 404-691, 404-704, 404-865, 404-927, 537-1223, 669-1195, 794-908, 930-978, 1063-1111
100/6046406CB1/ 549	1-548, 1-549, 34-549
101/6743529CB1/ 520	1-229, 3-520, 42-520, 52-157, 85-520, 89-520
102/7283809CB1/ 950	1-347, 1-480, 1-485, 1-501, 1-541, 1-746, 1-926, 91-950, 347-927, 383-950, 394-927, 413-930, 427-927, 428-927, 438-927, 471-949
103/7637563CB1/ 913	1-520, 1-589, 1-630, 172-913
104/7663814CB1/ 640	1-640, 57-561
105/8001939CB1/ 1113	1-587, 1-617, 18-612, 18-617, 175-393, 452-1082, 505-1113, 666-716, 824-1113
106/8191019CB1/ 933	1-440, 2-442, 3-442, 294-615, 294-762, 294-811, 295-811, 297-769, 345-811, 372-933, 457-811, 480-811, 514-811
107/919788CB1/ 1280	1-638, 13-937, 44-639, 82-662, 222-792, 236-790, 239-793, 240-792, 288-787, 290-792, 333-640, 362-792, 374-627, 383-659, 387-792, 394-781, 401-662, 409-793, 418-694, 419-675, 451-615, 451-617, 451-619, 451-646, 451-651,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	451-661, 451-705, 451-718, 451-719, 451-734, 451-742, 451-749, 451-753, 451-773, 451-785, 451-792, 451-812, 451-833, 451-841, 451-861, 451-868, 451-874, 451-1029, 451-1048, 452-623, 452-640, 454-712, 466-1176,
	467-743, 473-793, 490-793, 494-938, 499-792, 513-1145, 516-773, 516-791, 520-1280, 529-712, 532-1212, 547-719, 568-1015, 589-682, 589-793, 589-1010, 704-894
108/4758058CB1/ 697	1-555, 1-682, 37-555, 365-527, 365-534, 384-521, 384-522, 393-697
109/7499835CB1/ 723	1-174, 20-263, 27-279, 28-285, 28-294, 30-239, 30-259, 30-294, 31-243, 34-275, 36-251, 36-270, 36-273, 37-226, 37-253, 37-254, 52-581, 63-249, 63-279, 71-289, 71-293, 73-291, 74-292, 75-264, 75-268, 75-294, 81-289, 87-294,
	88-260, 89-294, 90-264, 90-276, 90-294, 91-294, 92-285, 95-268, 96-251, 214-702, 257-472, 285-723, 295-699, 295-706, 295-720, 303-525, 313-550, 315-723, 317-546, 317-549, 325-565, 333-514, 341-609, 347-703, 350-545,
	363-565, 363-630, 364-580, 364-583, 364-589, 364-592, 365-658, 372-590, 376-655, 381-608, 410-717, 422-717, 427-649, 427-658, 427-664, 457-634, 474-630, 475-691
110/2484647CB1/ 1049	1-663, 8-551, 8-1049, 93-673, 189-907, 262-618, 267-748, 285-543, 302-687, 302-705, 302-748, 302-815, 307-787, 322-798, 323-592, 323-721, 324-581, 324-593, 324-642, 324-655, 324-688, 329-882, 338-809, 354-900, 365-614,
	373-750, 375-748, 376-652, 377-635, 396-593, 400-795, 437-702, 453-907, 454-898, 475-751, 476-764, 483-907, 492-781, 499-642, 529-907, 535-824, 542-648, 558-1041, 566-798, 567-814, 580-845, 587-845, 587-907,
	588-815, 598-845, 605-905, 612-748, 618-905, 636-907, 645-748, 650-730, 665-882, 671-748
111/2587034CB1/ 360	1-360, 97-220, 128-333
112/2702991CB1/ 1466	1-246, 1-432, 1-450, 1-507, 1-522, 1-523, 1-527, 1-553, 1-570, 1-578, 1-583, 1-609, 1-618, 1-621, 1-660, 1-673, 1-684, 1-688, 1-696, 1-725, 1-726, 1-818, 1-856, 149-624, 152-961, 192-692, 230-475, 230-510, 235-745, 253-726, 259-483,
	261-591, 288-662, 328-526, 335-1081, 412-1081, 423-658, 425-1006, 428-869, 459-1081, 479-1083, 508-1073, 663-1236, 663-1240, 663-1270, 670-1270, 714-1279, 734-1244, 739-1282, 825-1302, 846-1413, 851-1393, 868-1424,
	871-1466, 873-1210, 898-1427, 905-1192, 906-1461, 919-1192, 942-1203, 958-1202, 968-1253, 976-1237, 1018-1439, 1029-1296, 1034-1309, 1047-1343, 1058-1273, 1060-1307, 1112-1150, 1112-1151, 1114-1285, 1148-1187

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
113/2744736CB1/ 1724	1-233, 1-253, 1-384, 1-388, 1-467, 1-477, 1-487, 1-538, 1-592, 1-594, 1-596, 1-645, 1-650, 1-791, 12-844, 118-828, 197-433, 210-658, 214-658, 222-813, 222-963, 231-712, 341-517, 359-671, 444-949, 446-931, 474-1034, 479-948, 496-887,
	496-974, 503-918, 509-1243, 511-1156, 536-1147, 539-854, 548-1023, 556-956, 558-944, 606-1407, 629-1198, 649-784, 653-927, 704-1338, 715-1128, 755-1060, 755-1188, 756-1354, 776-1393, 794-953, 805-1388, 821-1231,
	857-1338, 882-1479, 887-1724, 892-1393, 949-1418, 974-1414, 986-1207, 996-1435, 1184-1311, 1255-1397
114/2915475CB1/ 778	1-284, 1-528, 1-579, 1-631, 1-740, 1-748, 1-760, 1-761, 1-778, 3-473, 3-550, 3-756, 3-758, 3-761, 60-538, 252-778, 254-774, 635-778, 701-778
115/3040427CB1/ 1974	1-271, 1-484, 1-511, 1-521, 1-532, 1-590, 19-617, 34-663, 86-721, 116-600, 265-716, 333-891, 362-634, 363-897, 366-1033, 370-1052, 374-1008, 376-1066, 379-1037, 383-614, 385-683, 385-972, 412-645, 412-674, 412-677, 412-843, 412-981, 412-985, 414-518, 416-890, 423-1046, 446-821, 465-637, 470-770, 471-679, 473-696, 474-1052, 483-868, 489-821, 489-905, 497-706, 497-744, 498-947, 498-1036, 500-950, 501-686, 501-763, 519-870, 526-902, 527-1047, 531-970, 557-1052, 565-962, 567-962, 572-1036, 574-1075, 581-1034, 588-962, 589-638, 589-905, 592-1036, 593-1052, 594-1030, 594-1056, 605-1045, 606-845, 614-1036, 616-1047, 625-1036, 626-1036, 632-1052, 639-905, 642-1036, 651-1052, 652-1052, 656-905, 657-905, 668-1030, 691-1013, 693-1052, 697-850, 709-1044, 712-1052, 723-1052, 730-1052, 734-1030, 762-1036, 766-1052, 770-1030, 770-1052, 772-1052, 783-1050, 786-1036, 805-1052, 809-1052, 835-1052, 836-1052, 843-1116, 857-1052, 858-1052, 858-1178, 860-1052, 860-1218, 860-1448, 872-1052, 872-1158, 877-1052, 883-1052, 883-1146, 889-1045, 907-1403, 913-1384, 957-1027, 1080-1605, 1404-1974, 1706-1734, 1738-1766
116/7499722CB1/ 990	1-334, 1-990, 95-220, 128-449, 313-759, 398-757, 513-751, 595-990
117/6776909CB1/ 951	1-555, 1-597, 1-951, 351-951
118/7280438CB1/ 1106	1-544, 1-553, 1-1106, 271-980, 341-980, 344-980, 408-1101, 521-980

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
119/7499809CB1/ 2026	1-376, 1-503, 25-184, 25-190, 36-210, 38-1300, 208-439, 208-773, 208-783, 208-840, 208-927, 208-1002, 211-753, 217-761, 233-594, 234-808, 238-485, 238-499, 259-548, 260-411, 271-872, 273-573, 273-721, 273-831, 283-463, 283-857, 289-868, 316-1018, 322-778, 343-1010, 356-960, 356-1004, 373-810, 384-701, 388-972, 410-673, 423-965, 447-979, 450-1094, 452-721, 468-1143, 469-1143, 472-830, 475-770, 476-1092, 480-1112, 484-764, 494-940, 501-1143, 513-1059, 515-1047, 516-787, 519-651, 538-1147, 545-1116, 547-1152, 547-1188, 555-1143, 555-1151, 560-800, 560-1075, 570-848, 570-982, 571-770, 571-1050, 572-1092, 596-836, 599-1032, 606-1018, 632-1221, 633-1150, 636-849, 672-922, 684-1191, 696-975, 696-1221, 703-1094, 717-1057, 729-968, 766-1098, 783-1158, 783-1159, 783-1221, 784-1221, 789-1087, 790-985, 790-1087, 791-1087, 807-1000, 877-1154, 880-1134, 922-1184, 932-1219, 953-1221, 956-1221, 964-1175, 1117-1224, 1187-1394, 1224-1898, 1388-1600, 1389-1424, 1389-1452, 1389-1472, 1389-1504, 1389-1512, 1389-1515, 1389-1521, 1389-1524, 1389-1529, 1389-1534, 1389-1538, 1389-1551, 1389-1559, 1389-1561, 1389-1581, 1389-1582, 1389-1587, 1389-1591, 1389-1593, 1389-1599, 1389-1600,
	1423-1600, 1435-1600, 1447-1600, 1454-1600, 1468-1600, 1480-1600, 1517-1600, 1534-1600, 1535-1595, 1540-1600, 1541-1600, 1542-1600, 1558-1600, 1598-1701, 1598-1736, 1598-1764, 1598-1810, 1598-1824, 1598-1828, 1598-1873,
	1598-1919, 1598-2011, 1598-2022, 1598-2026, 1608-1788, 1608-2026, 1614-2026, 1641-2026, 1647-2026, 1668-2026, 1681-2026, 1698-2026, 1709-2026, 1713-2026, 1735-1998, 1748-2026, 1772-2026, 1775-2026, 1777-2026,
	1793-2026, 1794-2026, 1797-2026, 1800-2026, 1801-2026, 1805-2026, 1809-2026, 1819-2026, 1826-2026, 1838-2026, 1840-2026, 1841-2026, 1844-1976, 1863-2026, 1870-2026, 1872-2026, 1880-2026, 1885-2026,
	1895-2026, 1896-2026, 1897-2026, 1921-2026, 1924-2026, 1931-2026, 1957-2026, 1958-2026, 1961-2026, 1997-2026
120/7499921CB1/ 2169	1-127, 1-197, 1-248, 1-273, 1-321, 1-363, 1-494, 1-666, 1-2169, 2-127, 4-522, 9-127, 9-243, 10-127, 12-127, 13-127, 14-127, 20-243, 21-127, 25-284, 50-127, 53-127, 104-350, 357-603, 520-930, 527-639, 539-682, 550-1154, 564-779, 574-934,
	593-932, 724-1126, 726-1310, 763-1349, 794-1431, 830-1351, 976-1154, 1005-1054, 1102-1360, 1105-1222, 1245-2169, 1285-1492, 1286-1510, 1290-1543, 1333-1705, 1344-1610, 1395-2169, 1418-1531, 1430-1952,
	1486-1999, 1513-1765, 1554-1852, 1577-2169, 1620-2010, 1695-1992, 1726-2015

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
121/2705858CB1/ 852	1-511, 1-568, 1-588, 2-852, 50-731, 50-740, 51-497, 51-537, 51-548, 51-625, 51-626, 51-628, 51-650, 51-657, 51-711, 51-740, 51-767, 51-799, 67-412, 97-654, 101-816, 130-485, 165-676, 173-565, 173-601, 173-680, 186-680, 210-680, 252-633, 252-712, 252-818
122/3069892CB1/ 1245	1-285, 1-543, 1-640, 1-655, 494-1244, 499-1239, 501-1225, 506-1223, 586-1245, 647-1243
123/3069586CB1/ 1924	1-630, 10-1924, 71-783, 146-795, 219-846, 249-530, 256-980, 303-597, 303-695, 309-647, 321-564, 356-744, 373-983, 377-806, 382-648, 419-842, 436-713, 455-800, 455-847, 482-1024, 503-763, 512-786, 700-1411,
	700-1428, 700-1454, 1030-1402, 1030-1410, 1031-1412, 1179-1676, 1179-1717, 1179-1767, 1179-1859, 1215-1743, 1249-1598, 1394-1679
124/7500104CB1/ 559	1-252, 1-275, 3-236, 3-512, 7-90, 21-559, 90-311, 90-447, 90-533, 90-541, 92-326, 93-329, 94-318, 97-326, 102-323, 112-516, 119-550, 124-347, 128-194, 128-386, 146-508, 154-396, 175-298, 176-547, 182-391, 183-436, 189-441, 192-385, 195-433,
	210-451, 213-471, 220-464, 230-429, 230-452, 232-471, 233-501, 236-272, 236-275, 236-278, 236-296, 236-334, 236-335, 236-339, 243-283, 243-284, 243-288, 243-339, 244-288, 244-481, 252-339, 257-339, 260-460, 276-339,
	276-539, 278-339, 287-339, 287-349, 287-369, 287-374, 287-377, 287-380, 287-390, 288-515, 288-538, 293-339, 295-326, 295-362, 295-390, 301-541, 301-547, 305-534, 308-536, 311-390, 327-390, 327-546, 329-559, 338-559,
	339-390, 341-559, 345-390, 346-377, 346-390, 353-390, 356-390, 359-390, 362-390
125/7500203CB1/ 653	1-268, 14-152, 14-224, 14-232, 14-251, 16-137, 16-269, 16-286, 18-105, 19-321, 21-250, 24-193, 24-279, 25-166, 25-177, 25-195, 25-218, 25-228, 25-300, 27-259, 27-281, 27-332, 28-149, 29-182, 29-208, 29-225,
	29-247, 29-256, 29-276, 31-653, 44-276, 52-245, 91-297, 91-338, 134-308, 134-320, 139-327, 242-520, 301-516
126/4843802CB1/ 1649	1-245, 1-352, 1-354, 1-371, 1-374, 1-589, 1-591, 1-621, 1-625, 1-701, 1-769, 1-866, 1-895, 1-1341, 5-603, 5-615, 6-627, 6-715, 88-847, 122-1015, 144-603, 375-868, 410-768, 418-669, 525-1303, 588-1225, 633-1407, 670-1355, 882-1013, 935-1342, 939-1649
127/5877522CB1/ 1255	1-568, 42-553, 89-713, 132-638, 154-290, 159-359, 259-836, 361-855, 529-1163, 549-1159, 556-836, 575-1158, 581-1150, 581-1177, 877-1255
128/617491CB1/ 1021	1-228, 1-573, 102-499, 108-202, 108-203, 108-546, 130-203, 131-202, 131-203, 131-549, 131-550, 131-681, 133-549, 138-443, 138-697, 168-549, 168-550, 168-567, 168-694, 168-802, 169-549, 169-738, 169-747, 186-811,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	247-716, 249-530, 302-949, 338-959, 369-978, 439-894, 473-991, 518-949, 551-1013, 552-1010, 552-1017, 553-1012, 556-1021, 637-883, 667-957, 672-997, 672-1010, 674-964, 685-990, 741-998, 856-951
129/6289901CB1/ 1167	1-252, 1-653, 1-1167, 37-447, 37-450, 189-834, 450-1167, 1013-1068
130/6817709CB1/ 1045	1-598, 37-564, 37-583, 37-697, 201-754, 202-665, 226-829, 231-895, 542-810, 612-1045, 918-1045
131/6849312CB1/ 762	1-571, 1-715, 1-762, 88-570
132/7409581CB1/ 1550	1-826, 53-842, 53-925, 62-748, 82-771, 82-779, 89-780, 185-656, 190-478, 191-465, 195-478, 227-931, 264-931, 292-478, 293-931, 314-931, 331-478, 345-478, 373-738, 400-478, 415-465, 415-478, 428-461, 432-478, 433-461, 544-931,
	553-1027, 571-931, 652-931, 809-977, 809-1067, 809-1070, 809-1133, 809-1196, 809-1310, 809-1358, 809-1372, 809-1434, 809-1438, 809-1550, 978-1040, 978-1116, 979-1144, 979-1164, 979-1165, 979-1233, 979-1268, 979-1269, 979-1273,
	989-1031, 1124-1162, 1124-1166, 1124-1169, 1131-1164, 1131-1166, 1351-1398, 1351-1436, 1351-1441, 1351-1489, 1352-1415, 1352-1482, 1352-1485, 1353-1397, 1361-1489, 1362-1407
133/7437113CB1/ 2803	1-720, 1-2488, 46-456, 86-710, 130-420, 130-544, 130-940, 141-803, 190-431, 193-736, 195-424, 266-709, 273-848, 323-574, 368-952, 454-998, 583-1101, 621-1163, 669-1218, 711-1332, 726-1244, 803-1261, 803-1280, 808-1166, 809-1144, 811-1308,
	813-1023, 827-1417, 828-1439, 888-1534, 895-1181, 907-1374, 938-1166, 958-1678, 1027-1157, 1128-1635, 1159-1763, 1162-1298, 1203-1534, 1266-1842, 1302-1854, 1320-1948, 1330-1901, 1330-1917, 1330-1924, 1351-1589, 1359-2012,
	1375-1912, 1386-1625, 1402-1766, 1406-1620, 1455-1907, 1457-2138, 1523-2074, 1526-1744, 1621-2075, 1629-2186, 1657-1910, 1730-2242, 1734-1963, 1738-2079, 1745-2365, 1802-2206, 1818-2360, 1823-2166, 1827-2086, 1830-2112,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1899-2645, 1914-2332, 1918-2391, 1970-2445, 1974-2213, 1982-2255, 1982-2446, 2053-2366, 2062-2330, 2143-2385, 2230-2505, 2300-2497, 2312-2446, 2312-2802, 2362-2803, 2422-2649, 2575-2738, 2583-2803, 2584-2794, 2587-2803
134/7500260CB1/ 627	1-154, 1-615, 134-597, 134-598, 134-599, 157-391, 157-600, 157-626, 157-627, 158-401, 159-627, 160-627, 161-426, 161-614, 163-627, 164-614, 174-627, 175-395, 176-619, 181-598, 183-620, 192-608, 197-616, 212-614, 215-627, 216-615, 217-627,
	218-612, 219-614, 220-627, 227-427, 240-614, 255-598, 255-609, 255-615, 255-616, 256-549, 260-627, 261-521, 262-614, 266-614, 266-617, 270-627, 275-616, 279-618, 285-614, 285-621, 320-610, 323-617, 341-614, 345-614, 350-605,
	350-618, 350-627, 357-589, 357-614, 358-613, 362-544, 362-546, 364-549, 368-627, 369-627, 391-614, 396-627, 404-597, 413-598, 421-616, 433-613, 433-616, 433-627, 450-614, 513-617
135/7659504CB1/ 2337	1-690, 6-1959, 61-690, 173-686, 366-690, 386-690, 394-837, 404-689, 429-453, 484-732, 540-814, 540-825, 540-898, 540-943, 540-946, 540-1031, 540-1050, 540-1058, 540-1085, 540-1088, 540-1198, 540-1204, 636-1403, 648-1295, 713-737,
	772-954, 902-944, 902-1035, 902-1114, 902-1196, 902-1318, 902-1343, 913-1030, 913-1112, 913-1167, 913-1246, 922-1162, 972-1246, 984-1343, 986-1115, 986-1119, 986-1247, 986-1280, 986-1379, 986-1427, 1063-1628, 1081-1199,
	1081-1203, 1081-1335, 1081-1361, 1081-1414, 1136-1246, 1140-1203, 1173-1427, 1174-1266, 1174-1322, 1174-1414, 1179-1246, 1182-1246, 1210-1343, 1242-1944, 1258-1414, 1263-1343, 1268-1343, 1282-1680, 1330-1633, 1342-1633,
	1350-1681, 1369-1427, 1474-1721, 1595-1884, 1595-1950, 1595-2158, 1595-2170, 1595-2231, 1595-2249, 1595-2303, 1598-2252, 1598-2261, 1598-2337, 1600-2258, 1690-2137, 1710-2018, 1814-1960, 1814-2084
136/821165CB1/957	1-399, 1-523, 1-567, 1-813, 2-258, 2-365, 2-511, 2-523, 154-784, 176-557, 258-521, 271-760, 274-830, 332-557, 584-950, 584-957
137/7499672CB1/ 1731	1-648, 21-1347, 52-82, 80-347, 133-730, 133-798, 204-614, 251-924, 331-1077, 491-780, 616-736, 616-769, 624-894, 642-1088, 741-1025, 747-1197, 755-879, 759-1029, 759-1033, 759-1047, 852-1075, 852-1294, 854-1328,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	888-1065, 891-1126, 903-1342, 904-1084, 1045-1346, 1053-1347, 1062-1232, 1066-1347, 1073-1346, 1078-1345, 1115-1250, 1133-1347, 1214-1346, 1276-1581, 1276-1731
138/7500276CB1/ 695	1-91, 1-98, 1-179, 1-203, 1-233, 1-254, 1-259, 1-459, 6-124, 35-233, 35-500, 109-361, 118-332, 121-492, 123-361, 132-355, 132-392, 133-323, 134-370, 136-367, 147-413, 153-387, 166-286, 184-400, 194-349, 196-695, 229-518, 254-633
139/1440723CB1/ 1468	1-1468, 41-553, 42-693, 371-825, 601-868, 632-1231, 632-1285, 637-1408, 759-1163, 879-1457, 1071-1145, 1071-1160, 1071-1191, 1071-1200, 1071-1330, 1071-1351, 1071-1397, 1077-1251, 1077-1329, 1082-1202, 1082-1255, 1085-1146,
	1090-1182, 1094-1271, 1139-1229, 1139-1271, 1139-1352, 1139-1441, 1139-1457, 1139-1466, 1214-1299, 1214-1366, 1214-1421, 1214-1459, 1214-1466, 1223-1461, 1288-1435, 1288-1461, 1292-1461, 1296-1369, 1345-1438, 1345-1462, 1345-1466
140/7479612CB1/ 708	1-386, 1-449, 1-454, 225-446, 225-708, 461-708
141/1391514CB1/ 1781	1-458, 1-665, 87-385, 87-398, 379-662, 379-905, 379-1037, 379-1039, 379-1050, 379-1186, 417-1202, 666-944, 923-1781
142/2102578CB1/ 1032	1-621, 9-258, 9-310, 27-682, 61-358, 262-757, 262-943, 263-712, 264-539, 315-764, 316-726, 316-807, 316-890, 316-964, 316-1010, 316-1028, 318-570, 341-716, 357-613, 357-914, 359-622, 368-714, 457-1032
143/3213122CB1/ 2870	1-261, 1-429, 1-513, 1-596, 1-2833, 17-567, 288-566, 320-1034, 337-645, 353-587, 413-993, 416-1134, 423-1267, 479-1045, 554-1086, 657-1335, 660-1283, 711-1319, 714-1320, 717-887, 722-1458, 738-1327, 766-1329, 769-1242, 788-1478,
	881-1466, 888-1473, 920-1475, 937-1461, 972-1613, 974-1483, 975-1475, 975-1493, 991-1490, 992-1532, 996-1510, 1071-1580, 1072-1259, 1094-1783, 1129-1653, 1143-1734, 1157-1797, 1163-1771, 1175-1718, 1196-1801, 1200-1757, 1214-1844,
	1283-1950, 1319-1965, 1348-1912, 1353-2100, 1365-2067, 1378-1884, 1411-1916, 1411-2001, 1429-1994, 1430-1883, 1443-2001, 1482-2099, 1493-2014, 1499-2289, 1520-2068, 1543-1654, 1544-2261, 1549-2231, 1585-2575, 1590-2271,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1629-1905, 1666-2570, 1671-2372, 1695-2328, 1726-2247, 1737-2396, 1755-2570, 1782-2461, 1805-1946, 1812-2433, 1841-2572, 1845-2554, 1894-2594, 1913-2378, 1936-2570, 1951-2555, 2046-2677, 2088-2570, 2095-2629,
	2200-2829, 2219-2862, 2260-2570, 2289-2591, 2295-2722, 2302-2870, 2367-2783, 2408-2804, 2456-2868
144/4326307CB1/ 2337	1-399, 137-724, 149-685, 149-829, 228-905, 423-509, 423-513, 423-526, 423-557, 423-629, 423-675, 423-732, 423-781, 597-875, 801-1308, 850-1333, 882-1204, 1082-1324, 1082-1497, 1082-1515, 1082-1660, 1082-1710, 1151-1707, 1220-1750,
	1240-1869, 1261-1829, 1318-1860, 1366-1581, 1400-1808, 1428-1873, 1455-2068, 1494-2081, 1499-1898, 1538-2112, 1555-1811, 1581-2184, 1622-1785, 1649-2288, 1691-2337, 1696-1976, 1698-2064, 1748-2337, 1752-1863,
	1759-2000, 1770-2184, 1860-2231, 1978-2241, 1978-2280, 1978-2337, 2152-2337
145/6037749CB1/ 728	1-483, 1-596, 1-728
146/6285519CB1/ 1952	1-706, 68-879, 89-1920, 97-948, 120-838, 124-961, 127-969, 141-734, 141-980, 149-825, 151-841, 162-942, 200-431, 237-626, 243-310, 258-310, 269-866, 289-341, 289-427, 289-942, 364-428, 452-1377, 489-1004, 578-1230, 581-1443, 656-1294,
	662-1163, 662-1608, 689-1397, 714-1264, 727-1627, 777-1293, 787-1551, 825-1444, 829-1504, 885-1652, 891-1553, 901-1551, 903-1650, 919-1624, 919-1634, 945-1765, 1000-1728, 1022-1773, 1028-1474, 1033-1920, 1038-1660, 1040-1672, 1077-1920,
	1092-1802, 1108-1898, 1127-1946, 1175-1936, 1194-1920, 1230-1920, 1257-1920, 1259-1920, 1263-1920, 1263-1936, 1264-1952, 1265-1923, 1277-1920, 1277-1925, 1277-1936, 1278-1920, 1281-1936, 1285-1920, 1287-1920, 1292-1920, 1295-1920,
	1305-1920, 1316-1940, 1319-1916, 1323-1920, 1355-1923, 1362-1920, 1371-1920, 1387-1920, 1449-1935, 1449-1936, 1535-1852, 1540-1920, 1587-1920, 1618-1920, 1624-1920, 1679-1920, 1686-1920

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
147/70336045CB1/ 2490	1-589, 11-248, 12-801, 30-157, 39-287, 39-549, 40-152, 41-156, 41-425, 43-295, 43-341, 45-154, 49-681, 73-795, 81-531, 125-773, 137-801, 172-385, 185-445, 192-566, 202-469, 242-874, 259-800, 274-759, 279-590, 285-685, 293-888, 295-559, 300-658, 312-795, 333-579, 334-591, 342-575, 342-611, 342-638, 347-577, 352-659, 358-631, 362-647, 369-914, 378-807, 385-562, 397-649, 397-655, 401-916, 402-765, 403-686, 405-902, 423-705, 435-543, 438-695, 438-706, 440-742, 442-704, 452-923, 458-1123, 463-706, 471-701, 477-805, 477-1040, 489-996, 496-770, 496-799, 498-763, 501-1015, 510-916, 511-909, 518-1070, 520-901, 527-679, 528-1200, 533-915, 536-763, 539-1128, 540-920, 557-1143, 558-844, 559-740, 559-916, 565-924, 570-940, 570-1046, 573-818, 577-829, 594-871, 595-916, 601-1037, 611-1152, 618-928, 619-851, 627-819, 627-856, 627-876, 634-909, 662-872, 662-908, 665-880, 667-883, 667-887, 687-897, 698-904, 698-910, 698-916, 720-946, 722-1212, 725-886, 725-953, 745-922, 745-983, 753-906, 753-1019, 754-910, 754-915, 766-909, 770-1064, 777-1223, 840-1086, 848-1004, 854-1103, 861-950, 879-1360, 889-1148, 890-1496, 895-1176, 919-1125, 948-1180, 958-1068, 962-1224, 968-1080, 973-1208, 973-1354,
	981-1269, 983-1252, 1009-1250, 1083-1590, 1122-1309, 1146-1395, 1147-1399, 1147-1616, 1160-1433, 1182-1475, 1183-1450, 1190-1558, 1231-1798, 1252-1765, 1260-1520, 1261-1363, 1262-1482, 1275-1547, 1275-1778, 1279-1502, 1283-1569, 1286-1514, 1295-1566, 1299-1898, 1300-1477, 1312-1616, 1314-1566, 1317-1550, 1317-1552, 1319-1762, 1320-1503, 1322-1556, 1325-1622, 1334-1411, 1336-1591, 1340-1549, 1352-1560, 1357-1635, 1363-1490, 1367-1639, 1376-1606, 1377-1644, 1378-1892, 1383-1652, 1403-1679, 1404-1664, 1406-1634, 1411-1628, 1411-1629, 1411-1661, 1414-1644, 1428-1669, 1428-1699, 1433-1785, 1437-1702, 1444-1572, 1448-1689, 1449-1720, 1450-1696, 1450-1700, 1450-1726, 1452-1712, 1455-1678, 1455-1739, 1457-1741, 1460-1727, 1479-1737, 1483-1609, 1483-2088, 1494-1738, 1497-1797, 1505-1801, 1512-1786, 1514-1781, 1519-1793, 1519-1806, 1520-1835, 1521-1726, 1521-1780, 1521-1834, 1540-1979, 1555-1662, 1555-1836, 1559-2198, 1562-1855, 1574-1864, 1593-1837, 1595-1855, 1596-1857, 1597-1818, 1601-1855, 1607-1849, 1611-1856, 1615-1815, 1615-1817, 1615-1855, 1615-1856, 1622-2025, 1628-1869, 1632-1861, 1636-1890, 1637-1906, 1640-1897, 1640-2008,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1641-1869, 1641-2255, 1645-1900, 1647-1953, 1648-1897, 1648-2218, 1651-2108, 1651-2157, 1668-1895, 1672-1925, 1672-1973, 1673-1932, 1674-1822, 1674-1887, 1682-1821, 1693-1930, 1704-1955, 1704-1962, 1718-1956, 1733-1953, 1733-1999, 1742-1975, 1745-2042, 1746-2062, 1747-1786, 1747-2005, 1748-1996, 1754-1982, 1754-2096, 1754-2147, 1755-2021, 1755-2309, 1770-1954, 1773-2058, 1777-2227, 1791-2049, 1791-2185, 1800-2100, 1805-2052, 1810-2030, 1832-2060, 1834-2017, 1844-2009, 1844-2101, 1849-2065, 1851-2017, 1852-2278, 1856-2156, 1856-2183, 1858-2144, 1860-2089, 1862-2276, 1868-2278, 1871-2101, 1883-2278, 1885-2183, 1891-2254, 1896-2155, 1897-2352, 1897-2368, 1902-2134, 1921-2154, 1927-2409, 1931-2160, 1933-2180, 1936-2168, 1936-2278, 1937-2162, 1948-2021, 1948-2110, 1948-2157, 1948-2227, 1948-2415, 1957-2278, 1958-2236, 1958-2277, 1982-2233, 1984-2246, 2012-2202, 2018-2278, 2020-2268, 2021-2269, 2023-2279, 2024-2270, 2024-2279, 2026-2272, 2050-2401, 2073-2273, 2084-2278, 2084-2308, 2088-2329, 2089-2242, 2091-2278, 2094-2245, 2094-2278, 2094-2340, 2094-2391, 2098-2326, 2100-2278, 2111-2374, 2124-2279, 2126-2327, 2131-2359, 2142-2279,
	2221-2275, 2279-2482, 2311-2490, 2344-2402
148/7153577CB1/ 1051	1-1051, 4-626, 197-626, 282-364, 555-1051, 651-690
149/7500299CB1/ 586	1-586, 50-458, 50-461, 72-461, 153-418, 153-586, 156-423, 178-453, 197-332, 199-453, 253-514, 261-409, 272-571, 274-495, 274-530, 301-579, 302-586, 307-540, 348-586, 444-586
150/7480218CB1/ 3110	1-762, 94-880, 549-1095, 596-719, 659-1312, 659-1325, 665-893, 700-1208, 761-934, 785-1447, 807-1614, 855-1619, 867-1380, 867-1540, 867-1614, 867-1666, 875-1159, 902-1552, 905-1131, 975-1692, 1048-1434, 1061-1253, 1110-1580, 1134-1276, 1134-1443, 1134-1451, 1197-1338, 1237-1517, 1237-1839, 1261-1446, 1321-1948, 1421-2109, 1470-1747, 1470-1751, 1470-1880, 1470-1987, 1500-2129, 1545-2169, 1565-1711, 1576-2015, 1581-1633, 1607-2196, 1609-1960, 1610-1868, 1637-2185, 1746-1980, 1746-2136, 1746-2187, 1765-2191, 1772-2059, 1817-2052, 1824-2187, 1828-2187, 1853-2482, 1938-2203, 1942-2184, 1947-2188, 1950-2275, 1965-2225, 1965-2478, 1992-2675, 1998-2324, 2022-2504, 2045-2853, 2067-2327, 2071-2693, 2116-2504, 2139-2280, 2161-2899, 2188-2899, 2190-2792, 2207-2468, 2209-2328, 2209-2406, 2209-2659, 2209-2759, 2225-2565, 2240-2888, 2240-2899, 2243-2812, 2253-2876, 2262-2510, 2267-2781, 2268-2856, 2270-2856, 2283-2887, 2286-2897, 2303-2442, 2313-2733, 2338-2888, 2354-2870, 2354-2899, 2383-2683, 2391-2658, 2396-2888, 2398-2733, 2410-2887, 2420-2882, 2421-2886, 2427-2482, 2437-2896, 2452-2896, 2456-2658, 2460-2888, 2461-2896, 2474-2733, 2484-2899,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	2497-2774, 2510-2899, 2516-2856, 2518-2883, 2518-2885, 2518-2896, 2520-2831, 2526-2879, 2527-2896, 2528-2896, 2529-2878, 2553-2897, 2553-3110, 2560-2899, 2567-2881, 2570-2839, 2608-2896, 2636-2899, 2655-2899, 2701-2899, 2702-2894
151/7501159CB1/ 1852	1-1127, 1045-1845, 1137-1635, 1546-1852
152/7501932CB1/ 2057	1-779, 5-779, 48-779, 59-779, 68-779, 89-779, 137-779, 156-779, 168-779, 212-779, 245-779, 251-779, 269-779, 296-779, 297-779, 393-775, 398-779, 416-915, 462-779, 467-779, 611-779, 623-1171, 705-1189, 720-1496, 732-1378, 871-1162,
	871-1423, 886-1475, 895-1383, 925-1578, 1021-1701, 1160-1854, 1167-1497, 1167-1854, 1167-1855, 1168-1577, 1170-1938, 1194-1637, 1239-1834, 1293-1904, 1295-1562, 1307-1929, 1345-1596, 1438-2057, 1635-1895, 1710-1911, 1892-1966
153/7501111CB1/ 1848	1-215, 1-423, 1-519, 1-620, 3-215, 6-608, 6-692, 6-734, 6-761, 6-776, 7-771, 7-784, 8-405, 8-714, 13-647, 13-743, 13-749, 14-646, 14-689, 14-705, 14-719, 14-738, 14-745, 14-761, 14-778, 14-784, 14-850, 15-587, 15-620, 15-647, 15-701, 15-702,
	15-703, 15-718, 15-719, 15-720, 15-721, 15-728, 15-738, 15-741, 15-745, 15-748, 15-759, 15-771, 15-772, 15-773, 15-776, 15-778, 15-779, 15-784, 15-815, 15-850, 15-1848, 16-708, 16-730, 16-775, 17-738, 17-761, 18-257, 48-214, 111-561,
	129-784, 163-339, 201-756, 212-736, 237-682, 283-512, 302-719, 360-754, 848-878, 848-879, 848-899, 848-935, 848-964, 850-872, 851-932
154/7501113CB1/ 1616	1-424, 1-520, 1-621, 7-609, 7-693, 7-735, 7-762, 7-813, 8-790, 8-817, 9-406, 9-715, 14-648, 14-744, 14-750, 14-780, 15-647, 15-690, 15-706, 15-720, 15-739, 15-762, 15-784, 15-787, 15-795, 15-875, 16-588, 16-621, 16-648, 16-702,
	16-703, 16-704, 16-712, 16-719, 16-720, 16-721, 16-722, 16-739, 16-742, 16-746, 16-749, 16-760, 16-772, 16-773, 16-774, 16-786, 16-787, 16-789, 16-793, 16-804, 16-807, 16-808, 16-810, 16-816, 16-821, 16-822, 16-833, 16-872, 16-886,
	16-900, 16-901, 16-906, 16-907, 16-908, 16-1616, 17-709, 17-731, 17-776, 17-807, 17-929, 18-739, 18-795, 19-258, 31-932, 41-864, 49-215, 78-890, 112-562, 130-816, 164-340, 193-932, 202-757, 213-737, 238-683, 282-867, 284-513, 291-865, 294-927,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	303-720, 361-755, 363-860, 380-855, 400-865, 780-1556, 780-1558, 782-804, 782-1529, 793-1525, 882-1529, 931-1441, 932-1320, 933-1529, 988-1321, 1017-1320, 1076-1318, 1086-1321, 1103-1529, 1127-1321, 1184-1321, 1204-1529
155/7501118CB1/ 1568	1-215, 1-423, 1-519, 1-620, 2-207, 3-215, 6-608, 6-692, 6-696, 6-703, 7-707, 7-719, 8-405, 8-714, 13-647, 13-719, 13-721, 14-646, 14-689, 14-703, 14-705, 14-714, 14-719, 14-720, 14-721, 15-587, 15-620, 15-647, 15-701, 15-702, 15-703,
	15-711, 15-717, 15-718, 15-719, 15-720, 15-721, 15-734, 15-826, 15-1568, 16-708, 16-721, 17-721, 18-257, 48-214, 111-561, 163-339, 212-721, 237-682, 283-512, 302-719, 721-1513, 722-1481, 749-1469, 781-1512, 803-1481, 883-1393,
	884-1272, 885-1481, 940-1273, 969-1272, 1028-1270, 1055-1481, 1156-1481
156/7501128CB1/ 1799	1-757, 1-801, 1-836, 1-1798, 650-1382, 861-1594, 861-1628, 861-1742, 861-1795, 861-1799, 934-1754, 1269-1796, 1355-1607, 1488-1734, 1488-1739
157/7501920CB1/ 3395	1-229, 5-3395, 23-237, 24-229, 25-237, 28-899, 37-706, 37-883, 37-904, 88-1044, 252-1049, 260-477, 260-829, 260-831, 260-891, 267-684, 280-1076, 292-1252, 301-1190, 325-719, 327-824, 344-819, 364-804, 364-829, 379-872, 396-730, 450-1285, 458-1474, 463-1205, 471-1259, 476-1299, 550-1478, 551-1497, 551-1661, 557-1571, 569-1483, 589-1380, 589-1423, 589-1493, 597-1543, 604-1567, 610-1354, 615-1122, 627-1513, 629-1447, 629-1545, 631-1445, 631-1464, 636-1359, 665-1313, 685-1303, 685-1452, 690-1311, 692-1460, 703-1313, 704-1313, 706-1240, 716-1313, 717-1460, 726-1580, 744-1310, 744-1313, 746-1313, 748-1024, 752-1213, 771-1012, 771-1015, 771-1278, 771-1291, 771-1326, 771-1363, 771-1399, 771-1426, 771-1446, 771-1466, 771-1477, 771-1494, 771-1504, 771-1514, 771-1522, 771-1537, 771-1540, 771-1552, 771-1555, 771-1557, 771-1558, 771-1565, 771-1611, 771-1612, 771-1613, 771-1627, 771-1635, 771-1653, 771-1654, 771-1655, 771-1656, 771-1657, 771-1658, 771-1660, 771-1671, 771-1677, 771-1678, 771-1679, 771-1681, 771-1697, 771-1701, 771-1702, 771-1726, 772-1660, 773-1558, 778-1299, 793-1604, 836-1313, 844-1434, 844-1779, 860-1600, 861-1742, 861-1744, 937-1400,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	950-1308, 950-1311, 954-1313, 990-1442, 1007-1504, 1008-1308, 1008-1310, 1008-1311, 1008-1313, 1064-1947, 1169-2055, 1189-1791, 1212-1690, 1212-1869, 1215-1477, 1228-1400, 1294-2245, 1296-1984, 1296-2310, 1302-2160, 1319-2238, 1332-2307, 1343-2305, 1348-2307, 1352-2147, 1352-2189, 1354-2147, 1354-2155, 1355-2150, 1362-2307, 1369-1961, 1370-2225, 1374-2326, 1377-2307, 1380-2307, 1381-2307, 1383-1998, 1384-2307, 1386-2304, 1391-2207, 1391-2307, 1391-2337, 1392-2304, 1394-2307, 1396-2307, 1400-2307, 1402-2307, 1403-2310, 1404-2307, 1404-2379, 1405-2252, 1405-2307, 1406-2307, 1409-2307, 1414-2307, 1422-2307, 1424-2307, 1427-2307, 1429-2307, 1430-2307, 1432-2307, 1433-2307, 1436-2307, 1437-2216, 1437-2259, 1439-2287, 1444-2305, 1451-2233, 1451-2307, 1460-2307, 1462-2307, 1463-2307, 1464-2307, 1467-2283, 1470-2401, 1489-1679, 1496-2307, 1497-2147, 1498-2229, 1502-2307, 1503-2307, 1505-2337, 1506-2307, 1531-2307, 1534-2307, 1534-2379, 1535-2336, 1536-2307, 1546-2510, 1567-2307, 1584-1997, 1590-1836, 1590-1841, 1590-2102, 1641-1971, 1659-2615, 1660-2604, 1663-1998, 1696-2534, 1698-2533, 1701-2461, 1701-2622, 1702-2536, 1730-2587, 1742-2307,
	1742-2630, 1762-2582, 1771-1998, 1801-2254, 1882-2690, 1885-2548, 1909-2432, 1927-2882, 1939-2882, 1997-2379, 1997-2549, 1998-2882, 2029-2884, 2033-2884, 2035-2884, 2038-2882, 2060-2449, 2074-2334, 2076-2882, 2092-2882, 2093-2882, 2094-2884,
	2095-2882, 2095-2884, 2096-2884, 2099-2882, 2100-2884, 2102-2882, 2114-2882, 2124-2882, 2127-2423, 2127-2886, 2130-2884, 2133-2882, 2138-2884, 2139-2882, 2140-2882, 2141-2882, 2142-3029, 2150-2882, 2151-2882, 2151-3046, 2153-2882,
	2156-2882, 2170-2882, 2174-3055, 2177-2851, 2201-2445, 2201-2882, 2209-2882, 2243-2882, 2254-2882, 2275-2550, 2310-3135, 2321-3135, 2335-3135, 2363-3130, 2363-3135, 2366-3004, 2372-3135, 2380-3223, 2383-2882, 2402-3224, 2411-3135,
	2413-3135, 2427-3224, 2435-3135, 2442-3223, 2449-3224, 2463-2832, 2476-3224, 2481-3220, 2484-3224, 2487-3224, 2494-2885, 2503-3224, 2504-2752, 2518-3223, 2544-3015, 2550-3015, 2593-3135, 2797-3224

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
158/7510325CB1/ 3567	1-642, 1-692, 1-803, 1-828, 1-832, 1-857, 1-3567, 113-328, 113-682, 113-742, 113-900, 118-535, 131-927, 142-1133, 143-1103, 144-1057, 152-1041, 176-570, 178-675, 195-670, 215-655, 215-680, 230-723, 247-581, 288-996, 301-1136, 309-1056, 309-1325, 318-1006, 319-1143, 322-1110, 327-1150, 358-933, 382-1065, 391-1099, 401-1329, 402-1348, 402-1512, 408-1422, 420-1334, 424-1251, 440-1231, 440-1274, 440-1344, 448-1394, 450-1278, 455-1418, 461-1205, 464-891, 466-973, 478-1364, 480-1298, 480-1396, 482-1277, 482-1296, 482-1315, 487-1210, 493-997, 516-1164, 536-1154, 536-1303, 541-1162, 543-1311, 554-1164, 555-1164, 557-1091, 567-1164, 568-1311, 577-1431, 595-1239, 597-1164, 599-1314, 603-1064, 622-863, 622-866, 622-1129, 622-1142, 622-1177, 622-1250, 622-1277, 622-1297, 622-1317, 622-1328, 622-1345, 622-1355, 622-1365, 622-1373, 622-1388, 622-1391, 622-1406, 622-1408, 622-1409, 622-1416, 622-1462, 622-1463, 622-1464, 622-1478, 622-1486, 622-1504, 622-1505, 622-1506, 622-1507, 622-1508, 622-1509, 622-1511, 622-1522, 622-1528, 622-1529, 622-1530, 622-1532, 622-1548, 622-1552, 622-1553, 622-1577, 623-1511, 624-1409, 629-1150, 644-1455, 659-1290, 687-1164,
	688-1276, 695-1565, 695-1630, 695-1650, 701-1485, 706-1462, 711-1451, 712-1593, 712-1595, 728-1397, 748-1166, 748-1192, 788-1251, 801-1159, 801-1162, 805-1164, 819-1690, 824-1513, 841-1293, 845-1738, 858-1355, 859-1159, 859-1161, 859-1162, 859-1164, 860-1576, 863-909, 908-1701, 915-1798, 922-1712, 933-1762, 1011-1758, 1020-1906, 1040-1642, 1063-1541, 1063-1630, 1079-1251, 1147-1835, 1147-2069, 1147-2161, 1153-1841, 1153-2011, 1160-2096, 1162-2138, 1170-2089, 1183-2158, 1194-2156, 1196-1528, 1199-1967, 1199-2158, 1203-1998, 1203-2040, 1205-1998, 1205-2006, 1206-2001, 1207-1973, 1208-2186, 1213-2158, 1214-1849, 1215-2069, 1221-2076, 1225-2177, 1228-2158, 1231-2158, 1232-2158, 1234-1849, 1235-2158, 1237-2155, 1242-2058, 1242-2158, 1242-2188, 1243-2155, 1245-2158, 1247-2158, 1251-2158, 1253-2158, 1254-2161, 1255-2158, 1255-2230, 1256-2103, 1256-2158, 1257-2049, 1257-2158, 1260-2158, 1265-2158, 1270-2175, 1273-2158, 1275-2158, 1276-2171, 1278-2158, 1279-1672, 1280-2158, 1281-2158, 1283-2158, 1284-2158, 1287-2158, 1288-2067, 1288-2110, 1290-2138, 1295-2156, 1302-2084, 1302-2158, 1311-2158, 1313-2158, 1314-2158, 1315-2158, 1318-2134,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1321-2252, 1336-1848, 1340-1530, 1342-1849, 1347-2158, 1348-1998, 1349-2080, 1353-2158, 1354-2158, 1356-2188, 1357-2158, 1367-1849, 1367-1997, 1373-1849, 1382-2158, 1385-2158, 1385-2230, 1386-2187, 1387-2158, 1397-2361, 1418-2158, 1435-1848, 1441-1687, 1441-1692, 1441-1953, 1444-1800, 1492-1822, 1510-2325, 1510-2466, 1511-2455, 1514-1849, 1547-2385, 1549-2384, 1551-1814, 1552-2312, 1552-2473, 1553-2387, 1579-1800, 1581-2438, 1582-1800, 1593-2158, 1593-2481, 1598-2110, 1613-2433, 1622-1849, 1638-1800, 1652-2105, 1674-2321, 1733-2541, 1736-2399, 1746-2581, 1760-2274, 1778-2733, 1790-2733, 1848-2230, 1848-2400, 1849-2733, 1865-2679, 1880-2735, 1882-2477, 1884-2735, 1886-2735, 1889-2733, 1905-2734, 1911-2300, 1925-2185, 1927-2733, 1943-2733, 1944-2733, 1945-2735, 1946-2733, 1946-2735, 1947-2735, 1950-2733, 1951-2735, 1953-2733, 1965-2733, 1975-2733, 1978-2274, 1978-2737, 1981-2735, 1984-2733, 1987-2733, 1989-2735, 1990-2733, 1991-2733, 1992-2733, 1992-2788, 1993-2880, 2001-2733, 2002-2733, 2002-2897, 2004-2733, 2005-2734, 2007-2733, 2008-2733, 2012-2680, 2014-2733, 2021-2733, 2025-2906, 2028-2702, 2028-2733, 2052-2296, 2052-2733,
	2060-2733, 2069-2733, 2072-2680, 2072-2767, 2094-2733, 2099-2733, 2105-2733, 2126-2401, 2133-2733, 2139-2733, 2141-2733, 2161-2986, 2165-2946, 2170-2947, 2171-2733, 2172-2986, 2174-2733, 2180-2733, 2186-2986, 2194-2729, 2212-2733,
	2214-2981, 2214-2986, 2215-2733, 2223-2986, 2231-3074, 2234-2733, 2236-2828, 2248-3075, 2253-3075, 2261-3071, 2262-2986, 2264-2986, 2274-2455, 2278-3075, 2282-3075, 2286-2986, 2293-3074, 2300-3075, 2327-3075, 2332-3071,
	2335-3075, 2338-3075, 2345-2736, 2354-3075, 2355-2603, 2369-3074, 2395-2866, 2401-2866, 2444-2986, 2453-3075, 2486-3075, 2536-3075, 2589-3075, 2622-3075, 2648-3075
159/7510966CB1 1906	1-1906, 176-570, 459-1414, 480-1280, 480-1281, 518-1413, 582-1413, 623-1414, 626-1413, 627-1413, 636-1410, 657-1414, 666-1414, 674-937, 766-1414, 815-1325, 817-1414, 820-1204, 825-1414, 872-1205, 875-1414, 895-1325, 901-1204,
	925-1559, 928-1414, 960-1202, 961-1414, 970-1205, 987-1414, 998-1072, 1011-1205, 1049-1452, 1068-1205, 1088-1414, 1112-1583, 1182-1855, 1184-1633, 1481-1901, 1485-1633, 1498-1901, 1532-1902
160/7386101CB1/ 2122	1-1887, 6-650, 98-374, 99-317, 99-366, 105-375, 369-486, 369-770, 374-750, 377-654, 377-827, 377-962, 377-982, 377-984, 377-985, 389-662, 389-807, 389-810, 398-818, 406-943, 451-1174, 451-1240, 451-1251, 618-1093, 678-819, 842-1063,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	917-1465, 1124-1391, 1339-1884, 1383-1998, 1411-1582, 1453-2038, 1485-1938, 1492-1679, 1492-2020, 1498-2052, 1520-2122, 1539-1740, 1604-2056, 1615-2088, 1620-1934, 1625-1896, 1632-2044, 1708-1954, 1721-1795, 1747-2008

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
81	1417062CB1	MONOTXN03
82	2007701CB1	TESTNOT03
83	2915695CB1	THYMFET03
84	2969449CB1	HEAONOT02
85	2994102CB1	KIDNFET02
86	3410251CB1	PROSTUS08
87	5330327CB1	SINTFER02
88	5532048CB1	BRAUTDR03
90	60129797CB1	BRSTTUT01
91	6246243CB1	TESTNOF01
92	6804755CB1	THYRDIE01
93	6856852CB1	BRAIFEN08
95	7493507CB1	THYRNOT03
96	3075994CB1	BONEUNT01
97	2378119CB1	BRAFNON02
98	2987418CB1	FIBPFEN06
99	4223862CB1	PANCNOT07
100	6046406CB1	BRABDIR02
101	6743529CB1	THP1NOT03
102	7283809CB1	BRAIFEJ01
103	7637563CB1	SEMVTDE01
104	7663814CB1	UTRSTME01
105	8001939CB1	LNODTUC02
106	8191019CB1	UTRSTMR02
107	919788CB1	KIDNNOT26
108	4758058CB1	NERDTDN03
109	7499835CB1	PANCNOT01
110	2484647CB1	THP1AZT01
111	2587034CB1	BRAITUT22
112	2702991CB1	BRSTTMT01
113	2744736CB1	PROSUNE04
114	2915475CB1	THYMFET03
115	3040427CB1	LSUBNOT03
117	6776909CB1	UTRSTMC01
118	7280438CB1	BMARTXE01
119	7499809CB1	IONCDPV07
120	7499921CB1	PROSNOT28
121	2705858CB1	PONSAZT01
122	3069892CB1	UTRSNOR01
123	3069586CB1	BRAHNOT02
124	7500104CB1	LPARNOT02
125	7500203CB1	PANCNOT01
126	4843802CB1	OSTENOT01
127	5877522CB1	BRAHNON05
128	617491CB1	PGANNOT01
129	6289901CB1	BRAUTDR03
130	6817709CB1	OVARDIJ01
131	6849312CB1	KIDNTMN03
132	7409581CB1	SKINBIT01

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
133	7437113CB1	ADRETUE02
134	7500260CB1	PROSBPS05
135	7659504CB1	PROSNON01
136	821165CB1	KERANOT02
137	7499672CB1	THYMNOT04
138	7500276CB1	LEUKNOT02
139	1440723CB1	SINTNOR01
140	7479612CB1	OVARTUT07
141	1391514CB1	BRAIFER06
142	2102578CB1	UTREDIT07
143	3213122CB1	BRABDIK02
144	4326307CB1	BRABNOE02
145	6037749CB1	PITUNOT06
146	6285519CB1	BRAHTDK01
147	70336045CB1	EOSIHET02
148	7153577CB1	BONEUNR01
149	7500299CB1	KERANOT02
150	7480218CB1	PANCTUT01
152	7501932CB1	OVARDIT06
154	7501113CB1	BRSTTUS08
155	7501118CB1	BRSTTUS08
156	7501128CB1	PENITUT01
157	7501920CB1	PENITUT01
158	7510325CB1	PENITUT01
159	7510966CB1	PROSTUT09
160	7386101CB1	PROSUNE04

Table 6

Library	Vector	Library Description
ADRETUE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from right adrenal tumor tissue removed from a 49-year-old Caucasian male during unilateral adrenalectomy. Pathology indicated adrenal cortical carcinoma comprising nearly the entire specimen. The tumor was attached to the adrenal gland which showed mild cortical atrophy. The tumor was encapsulated, being surrounded by a thin (1-3 mm) rim of connective tissue. The patient presented with adrenal cancer, abdominal pain, pyrexia of unknown origin, and deficiency anemia. Patient history included benign hypertension. Previous surgeries included adenotonsillectomy. Patient medications included aspirin, calcium, and iron. Family history included atherosclerotic coronary artery disease in the mother; cerebrovascular accident and atherosclerotic coronary artery disease in the father; and benign hypertension in the grandparent(s).
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BONEUNR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from an untreated MG-63 cell line derived from an osteosarcoma tumor removed from a 14-year-old Caucasian male (donor A) and using mRNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female (donor B) during an exploratory laparotomy and soft tissue excision. Pathology indicated giant cell tumor of the sacrum in donor B. Donor B's history included pelvic joint pain, constipation, urinary incontinence, unspecified abdominal/pelvic symptoms, and a pelvic soft tissue malignant neoplasm. Family history included prostate cancer in donor B.
BONEUNT01	pINCY	Library was constructed using RNA isolated from Saos-2, a primary osteogenic sarcoma cell line (ATCC HTB-85) derived from an 11-year-old Caucasian female.

Table 6

Library	Vector	Library Description
BRABDIK02	PSPORT1	This amplified and normalized library was constructed using pooled cDNA from three different donors. cDNA was generated using mRNA isolated from diseased vermis tissue removed from a 79-year-old Caucasian female (donor A) who died from pneumonia, an 83-year-old Caucasian male (donor B) who died from congestive heart failure, and an 87-year-old Caucasian female (donor C) who died from esophageal cancer. Pathology indicated severe Alzheimer's disease in donors A & B and moderate Alzheimer's disease in donor C. Patient history included glaucoma, pseudophakia, gastritis with gastrointestinal bleeding, peripheral vascular disease, chronic obstructive pulmonary disease, seizures, tobacco abuse in remission, and transitory ischemic attacks in donor A; Parkinson's disease and atherosclerosis in donor B; hypertension, coronary artery disease, cerebral vascular accident, and hypothyroidism in donor C. Family history included Alzheimer's disease in the mother and sibling(s) of donor A.
		Independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRABDIR02	pINCY	This random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day for 40 years).
BRABNOE02	PBK-CMV	This 5' biased random primed library was constructed using RNA isolated from vermis tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec.

Table 6

Library	Vector	Library Description
BRAFNON02	pINCY	This normalized frontal cortex tissue library was constructed from 10.6 million independent clones from a frontal cortex tissue library. Starting RNA was made from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Grossly, the brain regions examined and cranial nerves were unremarkable. No atherosclerosis of the major vessels was noted. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were also multiple small microscopic areas of cavitation with surrounding gliosis scattered throughout the cerebral cortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec.
		The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAHNON05	pINCY	This normalized hippocampus tissue library was constructed from 1.6 million independent clones from a hippocampus tissue library. Starting RNA was made from posterior hippocampus removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. The cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were small microscopic areas of cavitation with gliosis, scattered through the cerebral cortex. Patient history included cardiomyopathy, CHF, cardiomegaly, an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6

Library	Vector	Library Description
BRAHNOT02	pINCY	Library was constructed using RNA isolated from posterior hippocampus tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. The patient presented with sepsis, cholangitis, and post-operative atelectasis and pneumonia. Patient history included CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, congestive heart failure, hypothyroidism, and peripheral vascular disease. Previous surgeries included cholecystectomy and Bilroth I gastrectomy for ulcer. Patient medications included Lasix, Synthroid, Pancrease, Voltaren, Vicoden, Zantac and K-Dur.
BRAHTDK01	PSPORT1	This amplified and normalized library was constructed using pooled RNA isolated from archaecortex, anterior and posterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure.
		Previous surgeries included cholecystectomy and resection of 85% of the liver. 7.6x10e5 independent clones from this amplified library were normalized in 1 round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFEJ01	pRARE	This random primed 5' cap isolated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who died at 23 weeks' gestation from premature birth. Serologies were negative. Family history included diabetes in the mother.

Table 6

Library	Vector	Library Description
BRAIFEN08	pINCY	This normalized fetal brain tissue library was constructed from 400 thousand independent clones from a fetal brain tissue library. Starting RNA was made from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAITUT22	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the right frontal/parietal lobe of a 76-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a meningioma. Family history included senile dementia.
BRAUTDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled globus pallidus and substantia innominata tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTTMT01	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated recurrent grade 4, nuclear grade 3, ductal carcinoma. Angiolymphatic space invasion was identified. Left breast needle biopsy indicated grade 4 ductal adenocarcinoma. Paraffin embedded tissue was estrogen positive. Patient history included breast cancer and deficiency anemia. Family history included cervical cancer.
BRSTTUS08	pINCY	This subtracted library was constructed using 2.36M clones from a breast tumor library and was subjected to two rounds of subtraction hybridization with 2.32M clones from a prostate tissue library. RNA was isolated from breast tumor tissue removed from the right breast of a 46-year- old Caucasian female during a unilateral extended simple mastectomy with breast reconstruction. Pathology indicated an invasive grade 3 adenocarcinoma. Patient history included breast cancer. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. NAR (1991) 19:1954 and Bonaldo et al. Genome Research (1996) 6:791.

Table 6

Library	Vector	Library Description
BRSTTUT01	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated invasive grade 4 mammary adenocarcinoma of mixed lobular and ductal type, extensively involving the left breast. The tumor was identified in the deep dermis near the lactiferous ducts with extracapsular extension. Seven mid and low and five high axillary lymph nodes were positive for tumor. Proliferative fibrocysytic changes were characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Patient history included atrial tachycardia, blood in the stool, and a benign breast neoplasm. Family history included benign hypertension, atherosclerotic coronary artery disease, cerebrovascular disease, and depressive disorder.
EOSIHET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
FIBPFEN06	pINCY	The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a prostate fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round)reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation.
HEAONOT02	pINCY	Library was constructed using RNA isolated from aortic tissue removed from a 10-year-old Caucasian male, who died from anoxia.
IONCDPV07	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.

Table 6

Library	Vector	Library Description
KERANOT02	PSPORT1	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is a human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
KIDNFET02	pINCY	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
KIDNNOT26	pINCY	Library was constructed using RNA isolated from left kidney medulla and cortex tissue removed from a 53-year-old Caucasian female during a nephroureterectomy. Pathology for the associated tumor tissue indicated grade 2 renal cell carcinoma involving the lower pole of the kidney. Patient history included hyperlipidemia, cardiac dysrhythmia, metrorrhagia, normal delivery, cerebrovascular disease, atherosclerotic coronary artery disease, and tobacco abuse. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
KIDNTMN03	pINCY	This normalized kidney tissue library was constructed from 2.08 million independent clones from a pool of two libraries from two different donors. Starting RNA was made from right kidney tissue removed from an 8-year-old Caucasian female (donor A) who died from a motor vehicle accident and left kidney medulla and cortex tissue removed from a 53-year-old Caucasian female (donor B) during a nephroureterectomy. In donor B, pathology for the matched tumor tissue indicated grade 2 renal cell carcinoma involving the lower pole of the kidney. Medical history included hyperlipidemia, cardiac dysrhythmia, metrorrhagia, normal delivery, cerebrovascular disease, and atherosclerotic coronary artery disease in donor B. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LEUKNOT02	pINCY	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).

Table 6

Library	Vector	Library Description
LNODTUC02	pINCY	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from pelvic lymph node tumor tissue removed from a 42-year-old Caucasian female (donor A) during regional lymph node excision and removal of a solitary ovary and from left axillary lymph node tumor tissue from another donor (donor B). For donor A, pathology indicated Hodgkin's disease, nodular sclerosing type. The cells were reactive for CD15 (Leu-MI). The patient presented with nodular lymphoma and unspecified abdominal and pelvic symptoms. Patient history included diabetes during pregnancy and normal delivery. Previous surgeries included bilateral breast implants, appendectomy, bilateral tubal destruction and dilation and curettage. Patient medications included methylprednisone, Cefclor, and Naproxen. Family history included atherosclerotic coronary artery disease in the father and alcohol abuse in remission in the sibling. For donor B, pathology indicated metastatic adenocarcinoma.
LPARNOT02	pINCY	Library was constructed using RNA isolated from tissue obtained from the left parotid (salivary) gland of a 70-year-old male with parotid cancer.
LSUBNOT03	pINCY	Library was constructed using RNA isolated from submandibular gland tissue obtained from a 68-year-old Caucasian male during a sialoadenectomy. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
MONOTXN03	pINCY	Normalized, treated monocyte tissue library was constructed from 7.6 million independent clones from a treated monocyte library. Starting RNA was made from RNA isolated from treated monocytes from peripheral blood obtained from a 42-year-old female. The cells were treated with anti-interleukin-10 (anti-IL-10) and lipopolysaccharide (LPS). The anti-IL-10 was added at time 0 at 10 ng/ ml and LPS was added at 1 hour at 5ng/ml. The monocytes were isolated from buffy coat by adherence to plastic. Incubation time was 24 hours. cDNA synthesis was initiated using a NotI-anchored oligo(dT) primer. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996 6):791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The libraries were then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized, treated monocyte tissue libraries following soft agar transformation.
		The DNA was linearized with NotI and insert containing clones were size-selected by agarose gel electrophoresis and recircularized by ligation.

Table 6

Library	Vector	Library Description
NERDTDN03	pINCY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone.
		The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
OSTENOT01	pINCY	Library was constructed using RNA isolated from untreated osteoblasts removed from the clavicle of a 40-year-old male.
OVARDIJ01	pIGEN	This random primed 5' cap isolated library was constructed using RNA isolated from diseased right ovary tissue removed from a 47-year-old Caucasian female during total abdominal hysterectomy, dilation and curettage, bilateral salpingo-oophorectomy, repair of ureter, and incidental appendectomy. Pathology indicated endometriosis. Pathology for the associated tumor tissue indicated multiple leiomyomata. The left ovary contained a corpus luteum. There was endometriosis involving the posterior serosa. The patient presented with metrorrhagia and a benign neoplasm of the ovary. Patient history included normal delivery, joint pain in multiple joints, and unilateral congenital hip dislocation. Previous surgeries included total hip replacement. Patient medications included calcium. Family history included kidney cancer in the mother; atherosclerotic coronary artery disease and aortocoronary bypass of 3 coronary arteries in the father; benign hypertension and Hodgkin's disease in the sibling(s); and benign hypertension and cerebrovascular accident in the grandparent(s).

Table 6

Library	Vector	Library Description
OVARDIT06	pINCY	The library was constructed using RNA isolated from diseased left ovarian tissue removed from a 24-year-old Caucasian female during left ovary lesion excision. Pathology indicated endometriosis (endometrioma) of the left ovary, consisting of a tan-maroon collapsed cyst. The serosal surface was tan-purple and irregular with fibrous and fibrinous adhesions. The internal surface was maroon-green and ulcerated with no papillary excrescences. Microscopic sections revealed fragments of ovarian stroma associated with focal areas with numerous macrophages containing hemosiderin pigment, fibrosis and rare glands surrounded by endometrial stroma. The patient presented with pain, dysmenorrhea, and a pelvic mass.
OVARTUT07	pINCY	Library was constructed using RNA isolated from right ovarian tumor tissue removed from a 58-year-old Caucasian female during bilateral salpingo-oophorectomy, regional lymph node excision, destruction of peritoneal tissue, cystocele repair, and skin repair. Pathology indicated FIGO (International Federation of Gynecology and Obstetrics) grade 3 adenocarcinoma, serous type, forming a mass and entirely replacing the right ovary. The left pelvic sidewall revealed a microscopic focus of metastatic adenocarcinoma. Patient history included hyperlipidemia, thrombophlebitis, and carcinoma in situ of the cervix uteri. Family history included cerebrovascular disease, breast cancer, hyperlipidemia, atherosclerotic coronary artery disease, and heart failure.
PANCNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the pancreatic tissue of a 29-year-old Caucasian male who died from head trauma.
PANCNOT07	pINCY	Library was constructed using RNA isolated from the pancreatic tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
PANCTUT01	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.

Table 6

Library	Vector	Library Description
PENITUT01	pINCY	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
PGANNOT01	PSPORT1	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.
PITUNOT06	pINCY	Library was constructed using RNA isolated from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia.
PONSAZT01	pINCY	Library was constructed using RNA isolated from diseased pons tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

Table 6

Library	Vector	Library Description
PROSBPS05	pINCY	This subtracted prostate tissue library was constructed using 4.48x10e5 clones from diseased prostate tissue and was subjected to two rounds of subtraction hybridization with 1.56 million clones from a breast tissue library. The starting library for subtraction was constructed using RNA isolated from diseased prostate tissue removed from a 70-year-old Caucasian male during a radical prostatectomy and closed prostatic biopsy. Pathology indicated benign prostatic hypertrophy. Pathology for the matched tumor tissue indicated adenocarcinoma. The patient presented with elevated prostate specific antigen and induration. Patient history included benign hypertension, gastrointestinal bleed, cardiac dysrhythmia, cardiac arrest, hyperlipidemia, alcohol abuse and fractured mandible. Previous surgeries included splenectomy, cholecystectomy and inguinal hernia repair. Patient medications included Verapamil and antacids. Family history included benign hypertension, myocardial infarction and coronary atherosclerosis in the mother; tobacco abuse and lung cancer in the father; tobacco abuse, cerebrovascular accident and lung cancer in the sibling(s).
		The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from nontumorous breast tissue from a different donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996): 791.
PROSNON01	PSPORT1	This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
PROSNOT28	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during a radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason grade 5+4. The patient presented with elevated prostate specific antigen (PSA). Family history included lung and breast cancer.

Table 6

Library	Vector	Library Description
PROSTUS08	pINCY	This subtracted prostate tumor library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a prostate tumor library . The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al. Genome Research (1996) 6:791.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
PROSUNE04	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated LNCaP cell line, derived from prostate carcinoma with metastasis to the left supraclavicular lymph nodes, removed from a 50-year-old Caucasian male (Schering).
SEMVTDE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from seminal vesicle tissue removed from a 63-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node excision. Pathology for the associated tumor tissue indicated Gleason grade 2+3 adenocarcinoma in the right side of the prostate. Adenofibromatous hyperplasia was present. The patient presented with prostate cancer, elevated prostate specific antigen and prostatic hyperplasia. Patient history included kidney calculus, extrinsic asthma, benign bowel neoplasm, backache, tremor, and tobacco abuse in remission. Previous surgeries included adenotonsillectomy. Patient medications included Ventolin and Vanceril. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and acute myocardial infarction in the father; and stomach cancer and extrinsic asthma in the grandparent(s).
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.

Table 6

Library	Vector	Library Description
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
TESTNOF01	PSPORT1	This 5' cap isolated full-length library was constructed using RNA isolated from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
THP1AZT01	pINCY	Library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).
THP1NOT03	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
THYMFET03	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a Caucasian male fetus.
THYMNOT04	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Caucasian male, who died from anoxia. Serologies were negative. The patient was not taking any medications.
THYRDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).

Table 6

Library	Vector	Library Description
THYRNOT03	pINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
UTREDIT07	pINCY	Library was constructed using RNA isolated from diseased endometrial tissue removed from a female during endometrial biopsy. Pathology indicated in phase endometrium with missing beta 3, Type II defects.
UTRSNOR01	pINCY	Library was constructed using RNA isolated from uterine endometrium tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the endometrium was secretory, and the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the associated tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, and paraplegia. Family history included benign hypertension, type II diabetes, and hyperlipidemia.
UTRSTMC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from uterus tissue removed from a 49-year-old Caucasian female (donor A) during vaginal hysterectomy and bilateral salpingo-oophorectomy and from uterus tissue removed from a 55-year-old Caucasian female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. For donor A, pathology indicated inactive endometrium and cervix with no diagnostic changes. Pathology for the matched tumor tissue indicated multiple (6) intramural leiomyomata. The patient presented with excessive menstruation, deficiency anemia, and dysmenorrhea. Patient history included abdominal pregnancy, headache, and chronic obstructive asthma. Previous surgeries included hemorrhoidectomy, knee ligament repair, and intranasal lesion destruction. Patient medications included Azmacort, Proventil, Trazadone, Zostrix HP, iron, Premarin, and vitamin C. Family history included alcohol abuse, atherosclerotic coronary artery disease, upper lobe lung cancer, and carotid endarterectomy in the father;

Table 6

Library	Vector	Library Description
		breast fibroadenosis in the sibling(s); and acute myocardial infarction, liver cancer, acute leukemia, and breast cancer (central) in the grandparent(s). For donor B, pathology indicated proliferative endometrium and unremarkable cervix. The patient presented with excessive menstruation, pelvic pain, uterine leiomyoma, and endometriosis. Patient history included hypothyroid, normal delivery, bladder dilation, irritable colon, and endometrial hyperplasia. Previous surgeries included adenotonsillectomy. Patient medications included Synthroid and vitamins. Family history included atherosclerotic coronary artery disease and malignant breast neoplasm in the mother; malignant colon neoplasm and aterial embolism in the father; and drug abuse in the sibling(s).
UTRSTME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from uterus tissue removed from a 49-year-old Caucasian female during vaginal hysterectomy and bilateral salpingo-oophorectomy. Pathology for the matched tumor tissue indicated multiple (6) intramural leiomyomata. The patient presented with excessive menstruation, deficiency anemia, and dysmenorrhea. Patient history included abdominal pregnancy, headache, and chronic obstructive asthma. Previous surgeries included hemorrhoidectomy, knee ligament repair, and intranasal lesion destruction. Patient medications included Azmacort, Proventil, Trazadone, Zostrix HP, iron, Premarin, and vitamin C. Family history included alcohol abuse, atherosclerotic coronary artery disease, upper lobe lung cancer, and carotid endarterectomy in the father; breast fibroadenosis in the sibling(s); and acute myocardial infarction, liver cancer, acute leukemia, and breast cancer (central) in the grandparent(s).
UTRSTMR02	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A) and using mRNA isolated from myometrium removed from a 45-year-old female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. In donor A, pathology indicated the endometrium was secretory phase. The cervix showed severe dysplasia (CIN III) focally involving the squamocolumnar junction at the 1, 6 and 7 o'clock positions. Mild koilocytotic dysplasia was also identified within the cervix. In donor B, pathology for the matched tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. Patient history included stress incontinence, extrinsic asthma without status asthmaticus and normal delivery in donor B. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease in donor B.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
153	7501111	3412087H1	SNP00120809	172	1481	T	T	G	noncoding	n/d	n/a	n/a	n/a
154	7501113	3412087H1	SNP00120809	172	1247	T	T	G	noncoding	n/d	n/a	n/a	n/a
155	7501118	3412087H1	SNP00120809	172	1199	T	T	G	noncoding	n/d	n/a	n/a	n/a
158	7510325	3412087H1	SNP00120809	172	2792	T	T	G	noncoding	n/d	n/a	n/a	n/a
158	7510325	5439262H1	SNP00072289	114	2468	T	T	C	noncoding	n/a	n/a	n/a	n/a
159	7510966	3412087H1	SNP00120809	172	1131	T	T	G	noncoding	n/d	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-16, SEQ ID NO:18-39, SEQ ID NO:41-43, SEQ ID NO:45-53, SEQ ID NO:55-56, SEQ ID NO:59-60, SEQ ID NO:68, SEQ ID NO:71-72, SEQ ID NO:74-75, and SEQ ID NO:77-80,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:44 and SEQ ID NO:73,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to the amino acid sequence of SEQ ID NO:67,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:69,

f) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:76,

g) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, and

h) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:81-160,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

19. A method for treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of SECP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of SECP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of SECP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-80, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-80.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains

nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

91. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:36.

92. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:37.

93. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:38.

94. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:39.

95. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:40.

96. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:41.

97. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:42.

98. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:43.

99. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:44.

100. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:45.

101. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:46.

102. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:47.

103. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:48.

104. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:49.

105. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:50.

106. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:51.

107. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:52.

108. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:53.

109. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:54.

110. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:55.

111. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:56.

112. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:57.

113. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:58.

114. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:59.

115. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:60.

116. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:61.

117. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:62.

118. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:63.

119. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:64.

120. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:65.

121. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:66.

122. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:67.

123. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:68.

124. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:69.

125. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:70.